Transparency’s Influence on Human-collective Interactions

KARINA A. ROUNDTREE, Oregon State University
JASON R. CODY, United States Military Academy
JENNIFER LEAF, H. ONAN DEMIREL, and JULIE A. ADAMS, Oregon State University

Collective robotic systems are biologically inspired and advantageous due to their apparent global intelligence and emergent behaviors. Many applications can benefit from the incorporation of collectives, including environmental monitoring, disaster response missions, and infrastructure support. Transparency research has primarily focused on how the design of the models, visualizations, and control mechanisms influence human-collective interactions. Traditionally most transparency research has evaluated one system design element. This article analyzed two models and visualizations to understand how the system design elements impacted human-collective interactions, to quantify which model and visualization combination provided the best transparency, and provide design guidance, based on remote supervision of collectives. The consensus decision-making and baseline models, as well as an individual collective entity and abstract visualizations, were analyzed for sequential best-of-n decision-making tasks involving four collectives, composed of 200 entities each. Both models and visualizations provided transparency and influenced human-collective interactions differently. No single combination provided the best transparency.

CCS Concepts: • Human-centered computing → Human computer interaction (HCI); User studies;

Additional Key Words and Phrases: Transparency, human-collective interactions, collective systems

ACM Reference format:
Karina A. Roundtree, Jason R. Cody, Jennifer Leaf, H. Onan Demirel, and Julie A. Adams. 2022. Transparency’s Influence on Human-collective Interactions. ACM Trans. Hum.-Robot Interact. 11, 2, Article 19 (March 2022), 48 pages.
https://doi.org/10.1145/3507470

1 INTRODUCTION

Few evaluations have investigated how transparency, the principle of providing easily exchangeable information to enhance comprehension [34], influences interactions and behaviors between human operators and robotic spatial swarms (six evaluations) or colonies (one evaluation). This article’s first objective expands the existing transparency literature by assessing how different models (i.e., algorithms) and visualizations influence human-collective interactions and behavior.
Collective robotic systems, which are composed of many simple individual entities, exhibit biological behaviors found in spatial swarms \[7\], colonies \[20\], or a combination of both \[37\]. Understanding the influence of system design elements, such as the models \[11\], visualizations \[33\], and operator control mechanisms on human-collective team interactions is necessary to ensure desired outcomes (e.g., high performance). Integrating transparency into the system design elements can mitigate poor operator behaviors, help attain meaningful and insightful information exchanges between the operator and collective, as well as improve the human-collective’s overall effectiveness.

Many of the existing transparency evaluations that have assessed human-collective interactions and behavior have only focused on the influence of one system design element (e.g., control mechanism). Implementing the best identified system design elements may not always yield optimal results, as the resulting human-collective system may become less transparent due to unanticipated and undesired operator behaviors. This article’s second objective determines whether using the combined best model and visualization, previously analyzed independently by the authors \[11, 33\], provides the best transparency. Understanding how design elements interact to influence human-collective behavior is necessary to quantify transparency properly.

The evaluated task was a sequential best-of-n decision making problem, similar to a bee colony searching for a new hive location. A subset of the colony fly to a nearby tree branch, where they wait while scout bees search for a new hive location \[37\]. The bees exhibit spatial swarm behaviors during the initial flight, similar to those found in schools of fish \[12\] and flocks of birds \[5\]. Scout bees identify possible hive locations and evaluate each option relative to ideal hive criteria. The scouts return to the waiting colony to begin a selection process (i.e., colony behavior) entailing debate and building consensus on the best hive location (i.e., best-of-n \[40\]). After completing a consensus decision-making process, the bees travel to the new location, transitioning from colony based behaviors back to spatial swarm behaviors.

The best-of-n problem translates to many real world domains, such as searching for the highest-valued information in an environment, conducting a broad search, such as across a megacity, after a natural disaster to identify areas with the highest likelihood of survivors who require rescuing, or searching for the most dangerous hot spots during massive wildland fires. Adding an operator, who may possess information that a collective does not, can reduce the time to make decisions and ensure improved overall performance. The operator’s ability to influence the collective’s behavior positively relies on transparent systems that enables the operator to perceive accurately the collective’s state, comprehend what it is doing, and plans to do in the future. Transparency provided to a human supervisor \[36\], was analyzed using two models \[10\], the authors’ sequential best-of-n decision making model and another that served as a baseline behavior model, as well as two visualizations, a traditional collective representation, the Individual Agents interface \[32\], and an abstract Collective interface \[10\]. Focusing on the combined model and visualization, unlike the prior authors’ evaluations that assessed one design element (e.g., model), are necessary when communication and interaction with remote collectives will only occur via an interface. Understanding how the combined model and visualization designs impact the operator’s ability to positively influence the collective’s decision-making process informs this article’s third contribution, design guidelines to achieve transparency in human-collective systems.

This article analyzes the evaluation results from a different perspective than the authors two prior analyses: model-based \[11\] and visualization-based \[33\]. The model evaluation investigated the performance of two best-of-n models and a baseline model, with and without a human operator. The visualization evaluation \[33\] investigated how two collective visualizations impacted operators using the model evaluation’s dominate best-of-n decision-making model with environmental
bias. The transparency assessment considered the impact on individuals with different capabilities, operator comprehension, usability, and human-collective performance.

2 RELATED WORK

Understanding how entities communicate and interact to influence individual entity and global collective state changes is necessary to ground collective system design. Further, understanding prior collective system design elements and their influence on human-collective interactions is also presented. Finally, understanding how factors that affect transparency, or are influenced by transparency, such as explainability, usability, and performance, influence the human-collective system, is necessary to inform design decisions.

2.1 Collective Behavior

Spatial swarm systems are inspired by self-organized social animals (e.g., fish schools) [7] and exhibit intelligent, emergent behaviors as a unit by responding to locally available information [35]. Spatial swarms rely on distributed, localized, and implicit communication [22, 38]. Basic rules of repulsion, attraction, and orientation enable individual spatial swarm entities to position themselves relative to neighboring entities [2, 13]. Robotic colony entities exhibit unique roles, such as foraging, which adapt over time to maintain consistent states in changing conditions [45]. Colonies share information in a centralized location, such as honeybees inside a nest [37]. Positive feedback loops support gaining a consensus to change the colony’s behavior [39] and negative feedback mitigates saturation issues, such as food source exhaustion [6]. Additional collective behavior details are available in sources, such as Cody et al. [11] and Roundtree et al. [33].

2.2 Collective System Transparency and Influence on Human-collective Interactions

Many existing transparency evaluations investigated how control mechanisms influenced human-spatial swarm interactions and behavior (e.g., References [25, 27]). Two mechanisms were used to control a spatial swarm foraging in simple and complex environments [27, 28]: selection, which influenced a selected subgroup; and beacon, which exerted influence on entities within a set range. Selection outperformed beacon and was optimal in complex environments [27]; however, as the spatial swarm size increased, beacon required less operator influence [28]. Leader, predator, and mediator control mechanisms were assessed with regard to spatial swarm manageable and performance [25]. Leaders attracted entities towards them, predators repelled entities, and mediators allowed the operator to mold and adapt the spatial swarm. Workload increased when using leaders, decreased with predators, and remained relatively stable with mediators. Operators using leaders gathered the spatial swarm entities together and guided them in a particular direction. Spatial subswarms emerged and were pushed in different directions when using predators. Strategically placed mediators resulted in lower workload, suggesting that it may be easier to use. The quantity and quality of operator influence was investigated to determine when it becomes detrimental to the human-spatial swarm performance [43]. Operators moved a spatial swarm around using a dispersion algorithm (high autonomy) and user-defined go-to points (low autonomy). Complex environments containing numerous obstacles and small passageways required operator influence; however, too much control never allowed the automation to operate, resulting in a performance decline. Two operator interaction strategies emerged: (1) allow the autonomous algorithm to control spatial swarm movement or (2) manually break the spatial swarm into subgroups and guide them to explore different map areas.

Operator influence and information reliability was assessed during a best-of-n decision-making task [3] that required operators to place beacons in the environment to attract colony support. The individual entities’ directions were presented using a radial display surrounding the hub. Low
operator influence resulted in high performance when reliable information was provided, while high influence was best when inaccurate or incomplete site information was provided. Additional human-colony system evaluations are needed to establish a broader understanding of control mechanism influence on human-colony interactions, especially with imperfect communication.

Four spatial swarm visualizations were assessed based on the operator’s ability to predict the spatial swarm’s future state [42]. The full information visualization showed each individual entities’ position and heading, the centroid/ellipse showed a bounding ellipse at the center of the spatial swarm, the minimum volume enclosing ellipse showed individual entities at the spatial swarm’s edge, and random condition clustering showed individual entities spaced throughout the spatial swarm. The full information and centroid/ellipse visualizations enabled the most accurate predictions when estimating spatial qualities, while the bounding ellipse was preferred during low bandwidth situations. A metacognition model that enabled individual entities to monitor changes in the spatial swarm’s state and a visualization that communicated spatial swarm status during a convoy mission were assessed using spatial, audio, and tactile cues [21]. The primary task required monitoring the spatial swarm and responding to visualization signals while performing a secondary robotic planning task. The visualization enabled 99.9% accuracy of signal detection and recognition.

Transparency embedded in a traditional individual collective entity visualization, and an abstract visualization, was evaluated by Roundtree et al. for a single human operator-collective team performing a sequential best-of-n decision-making task [33]. Transparency was assessed by understanding how the visualization impacted operators with different individual capabilities, their information comprehension, the interface’s usability, and the human-collective team’s performance. The abstract visualization provided better transparency compared to the traditional visualization, because it enabled operators with individual differences and capabilities to perform relatively the same and promoted higher human-collective performance. The same abstract visualization was evaluated by Cody et al. to assess how different models (two sequential best-of-n decision-making models and one baseline model) influenced performance with and without a human operator [11]. The sequential best-of-n decision-making model that compensated for environmental bias without an operator performed slower, but made 57% more accurate hard decisions compared the sequential best-of-n model that only assessed a target’s value. The environmental bias compensated model required less operator influence and achieved 25% higher accuracy for hard decisions. Further analysis is needed to determine how the combination of models and visualizations influence human-collective interactions and which system design elements promote better overall transparency.

2.3 Transparency Factors

Transparency is the principle of providing information that is easy to use in an exchange between human operators’ and collective robotic systems to promote comprehension of intent, roles (e.g., decision-maker versus information gatherer), interactions, performance, future plans, and reasoning processes [34]. The term principle describes the results of a process of identifying what factors affect and are influenced by transparency, why those factors are important, how the factors influence one another, and how to design a system to achieve transparency. A subset of human-collective transparency factors, in Figure 1, is used in this article to assess the influence of transparency for the combination of different models and the Individual Agents and Collective visualizations on human-collective interactions. More detailed information about transparency factors and their definitions are provided in Roundtree et al. [34]. Seven factors impact transparency directly, shown as the blue ovals. The three highest total degree (number of in degree + number of out degree) direct factors are explainability, usability, and performance (dark blue). Information and
understanding (light blue) are not high degree factors, because explainability uses information to communicate and promote understanding. Observable, the ability to be perceived, and directable, the operator’s ability to guide problem solving [8], are not high degree factors due to the low number of in and out connections. Explainability and usability, a multifaceted quality that influences the operator’s perception of a system, are used to implement transparency in the presented models and visualizations. Performance can be used to assess the models and visualizations influence by determining how well the human-collective team was able to produce an output when executing a task [1].

Many factors that impact transparency are embedded in the different models or visualizations indirectly, the yellow rectangles. The timing and quantity of information visualized, such as collective status, requires considering the operator’s capability limitations [16], system limitations, as well as task and environmental impacts, to be explainable [4]. Human-collective efficiency and effectiveness can be improved by enabling the operator and collective to control aspects of the decision-making process via the model or control mechanisms. Visualizing information, such as predicted collective states, in a clear manner that helps alleviate the time and effort an operator must exert when integrating information to draw conclusions [24] and justify actions [17] is a strategy to promote efficiency and effectiveness. Poor judgments may occur if the operator lacks an understanding of the model’s reliability, due to inadequate training prior to interacting with the system, or the model is not memorable. The visualization usability may also contribute to negative behavior by hindering the operator’s perception and comprehension of the collectives’ current and predicted future actions. Poor judgments and human-collective interactions may cause operator dissatisfaction. Models that are not designed to leverage the operator’s and collective’s strengths to achieve a task, visualizations that do not provide needed information, and control mechanisms that do not promote effective influence over collective behavior will hinder human-collective performance, lower operator situational awareness (Endsley’s [14] perception, comprehension, and projection SA), impact workload negatively, and potentially compromise the team’s safety.
Understanding the relationships between the direct and indirect factors, and their relationship to transparency, is needed to assess metrics that can quantify how the transparency embedded in different combinations of models and visualizations influence human-collective interactions.

3 HUMAN-COLLECTIVE TASK

A single operator supervised and assisted four collective robotic systems, each composed of 200 simulated Unmanned Aerial Vehicles, performing a sequential best-of-\(n\) decision-making task. The human-collective team’s decision-making task required each collective to identify and select the highest-valued target, from a finite set of \(n\) options \([11]\), within a 500 \(m\) range of the current hub, the collective hub moved to the selected target and initiated a second target selection decision.

The four collective hubs were visible at the start of each trial. Targets became visible as each was discovered by a collective’s entities. The target’s value was assessed by the collective’s entities, who returned to their respective hub to report the target’s location and value. The collectives were only allowed to discover and occupy targets within their search range, but some targets were within proximity of multiple hubs. A collective’s designated search area changed after establishing at a new hub site. The operator was to prevent multiple collectives from merging or moving to the same target. When a collective moved to a target, the hub moved to the target location, and the target was no longer visible to the operator or available to other collectives. The collective that moved its hub to a target’s location first, when two collectives were investigating the same target, established its hub, while the second collective returned to its prior hub location.

A sequential best-of-\(n\) decision-making model (\(M_2\)) adapted an existing model (\(M_1\)), which based decisions on the target’s quality (i.e., value) \([31]\). The \(M_2\) performed better than the \(M_1\), as reported by Cody et al. \([11]\] and is incorporated into this article’s analysis. Information exchanges between a collective’s entities were restricted to occur inside the hub. Episodic queuing cleared messages when the collective entities transitioned to different states, which resulted in more successful and faster decision completion. Interaction delay and interaction frequency were added as bias reduction methods to consider a target’s distance from the collective hub and increase interactions among the collectives’ entities. Interaction delay improved the success of choosing the ground truth best targets, and interaction frequency improved decision time. The baseline model (\(M_3\)) allowed the collective entities to search and investigate potential targets, but the operator was required to influence the consensus-building element and select the final target. More detailed information about the models is provided in Cody et al. \([11]\].

The interface control mechanisms allowed the operator to alter the collectives’ internal states, including their levels of autonomy, throughout the sequential best-of-\(n\) selection process. The collective’s entities were in one of four states. Uncommitted entities explored the environment searching for targets and were recruited by other entities while inside of the collective’s hub. Collective entities that favored a target reassessed the target’s value periodically and attempted to recruit the collective’s other entities within the hub to investigate the specified target. Collective entities committed to a particular target once a quorum of support was detected or after interacting with another committed entity. Executing entities moved from the collective’s current hub location to the selected target’s location. A collective operated at a high level of autonomy by executing actions associated with potential targets independently.

The static map provided ecological validity and emulated a task searching an urban environment for potential locations of interest. Understanding the environment’s topography is necessary for identifying what type of vehicles (e.g., air versus ground vehicles) will be most effective at completing a task, depending on the environmental conditions. The highest-valued targets were a bright opaque green, while lower-valued targets had a more translucent green color. Targets within the
collective’s 500 m search range had different colored outlines, depending on the collective’s state: explored, but not currently favored; explored and favored; and abandoned targets.

The collective entities began each trial by exploring the environment in an uncommitted state, which transitioned to favoring as targets were assessed and supported. The collective committed to a target when 30% of the collective (60 entities) favored it. The collective moved to the selected target’s location once 50% of the collective (100 entities) committed to the target. The number of collective entities in a particular state or supporting a target was provided via the collective hub and target information pop-up windows that appeared relative to the respective collective’s hub or target. The operator was able to move the information windows by dragging the pop-up display.

The Individual Agents (IA) interface (see example in Figure 2(a)), developed using the Unity Real-Time Development Platform, displayed each individual collective entities’ location, along with the respective hubs, discovered targets, and other associated information. The discovered targets were initially white and transitioned to green when at least two individual collective entities evaluated the target. The individual collective entities’ state information, uncommitted (yellow), favoring (green), committed (blue), and executing (blue and moving to a new hub location), was conveyed via individual collective entity color coding.

The Collective interface, developed using the QT Platform, provided an abstract visualization that does not represent individual collectives’ entities [10], as shown in Figure 2(b). The collectives were rectangles with four quadrants representing the collectives’ states (uncommitted (U), favoring (F), committed (C), and executing (X)) and used a brighter white quadrant for a larger number of collective entities in a particular state. Targets contained two sections, where the top green section represented the target’s value (brighter and more opaque the green, the higher the value) and a bottom blue section indicated the number of collective entities favoring a particular target (brighter and more opaque the blue, the higher the number of collective entities). The collective’s outline moved from the hub to the target’s location to indicate the hub’s transition to the selected target.

The operator influenced a collective’s state via a collective request (see Figure 2(c)). Communication from the operator with the collective’s entities occurred inside the hub to simulate limited
real-world communication. The commands were communicated to the specified hub. The investigate command increased a collective’s support for a specific target by transitioning uncommitted entities (5% of the population) to the favoring state. Additional support for the same target was achieved by reissuing the investigate command repeatedly. An abandon command reduced a collective’s support for a specific target by transitioning favoring entities to the uncommitted state and only needed to be issued once for the collective to ignore a specified target. A collective’s entities stopped exploring alternative targets and moved to the selected target when the decide command was issued, which was available only when at least 30% of the collective supported the specified target. An operator was unable to further influence a collective once the decide command was issued.

A collective assignments section logged the operator’s issued commands with respect to particular collectives and targets, as well as indicated if the command was active or completed. Once a collective reached a decision, all prior commands were removed from the collective assignments log. The operator was only able to cancel an abandon command. Illegal messages were displayed in the system messages area and occurred when an operator requested an invalid command, which aroused when the operator attempted to issue an investigate command for targets outside of the collective’s search region; abandon newly discovered targets that did not have an assigned value (white targets); and issue decide commands when less than 30% of the collective supported a target. Additional detailed information about the human-collective task, as well as the IA and Collective interfaces, is provided in Cody et al. [11] and Roundtree et al. [33].

4 EXPERIMENTAL DESIGN

The primary research question for the within-model and between-visualization analyses was which model and visualization combination achieved better transparency? Four secondary questions were developed to investigate how the model and visualization combinations impacted the highest degree direct transparency factors in Figure 1. The first research question (R1) focused on understanding how the model and visualization combination influenced the operator. Individual differences, such as spatial capability, will impact an operator’s ability to interact with the interface effectively and cause different responses (i.e., loss of situational awareness or more workload). A model and visualization combination that can aid operators with different capabilities is desired. The explainability factor was encompassed in R2, which explored whether the model and visualization combination promoted operator comprehension. Perception and comprehension of the visualized information are necessary to inform future actions. Understanding which model and visualization combination promoted better usability, R3, will aid designers in promoting effective transparency. The final research question, R4, assessed which model and visualization combination promoted better human-collective performance. An ideal human-collective system performs a task quickly and accurately.

The independent variables included the within model variable, M2 and M3, the between visualization variable, IA versus Collective interfaces, and the trial difficulty (overall, easy, and hard). Trials that had a larger number of high-valued targets in closer proximity to a collective’s hub were deemed easy, while hard trials placed high-valued targets further away from the hub. The independent variables were consistent for all the research questions, while the dependent variable details are embedded into the sections associated with each research question.

4.1 Experimental Procedure

The experimental procedure required participants to complete a demographic questionnaire and a Mental Rotationstest [41]. The IA participants completed an additional Working Memory Capacity assessment. Upon completing of the demographic questionnaire, participants received training for
their respective interface. Two practice sessions occurred prior to each trial to ensure familiarity with the underlying sequential best-of-n ($M_2$) and baseline ($M_3$) models. Both evaluations always completed the $M_2$ trial prior to the $M_3$ trial to alleviate any learning effects. The participants were instructed that the objective was to aid each collective robotic system in selecting and moving to the highest-valued target two sequential times. A trial began upon completing the practice session. Each trial had two components (one easy and one hard) of approximately 10 minutes each. The simulation environment reset between the components with 16 new (not initially visible) targets. The easy and hard trial orderings were randomly assigned and counterbalanced across the participants. The situational awareness (SA) probe questions \cite{10}, a secondary task, were asked beginning at 50 seconds into the trial and were repeated at one-minute increments. Six SA probes were asked during each trial component, or 12 per trial. The trial ended after eight decisions, two per collective, or if the trial length exceeded 10 minutes, after six decisions were completed. Decision times were not limited. A post-trial questionnaire was completed after each trial, and the post-experiment questionnaire was completed before the evaluation termination.

### 4.2 Participants

The demographic questionnaire recorded participants’ age, gender, education level, and weekly hours on a desktop or laptop (0, less than 3, 3–8, and more than 8). The Mental Rotation Assessment \cite{41} required judging three-dimensional object orientations to assess spatial reasoning within a scoring range of 0 (low) to 24 (high). Effective spatial reasoning will enable participants to understand what collectives are, how the collectives move in an environment, and where the collectives are located. Understanding collective movement is necessary for the participant to effectively interact with the collectives to successfully complete the best-of-n decision-making task. The Working Memory Capacity assessment, only completed by IA participants, evaluated higher-order cognitive task performance, such as comprehension and reasoning \cite{15}. Working memory capacity can influence the participant’s ability to: (1) understand the collectives’ respective behaviors that inform effective interactions with the collectives (primary task) to select the highest-valued target while (2) responding to the intermittent SA probe questions (secondary task). A reading span test assessed whether a sentence was accurate while recalling a series of letters interspersed between sentences.

Fourteen females and 19 males (33 total) completed the IA evaluation at Oregon State University. Five participants were excluded due to inconsistent methodology (1) and software failure (4). The mean weekly hours spent on a desktop or laptop was 3.79, with a standard deviation (SD) = 0.5, median = 4, minimum (min) = 2, and maximum (max) = 4. The Mental Rotation Assessment \cite{41} mean was 12.36 (SD = 5.85, median = 12, min = 3, and max = 24) and the Working Memory Capacity mean was 86.14 (SD = 9.73, median = 89.5, min = 59, and max = 98) \cite{32}.

Twenty-eight participants, 15 females and 13 males, from Vanderbilt University completed the Collective evaluation. The weekly hours spent on a desktop or laptop was slightly higher than the IA participants (mean = 3.86, SD = 0.45, median = 4, min = 2, and max = 4) and the Mental Rotations Assessment was slightly lower (mean = 10.93, SD = 5.58, median = 10, min = 1, and max = 24) \cite{32}.

### 4.3 Analysis

The mixed analysis is based on 56 participants from both evaluations. The first 12 decisions made per participant using each model were analyzed. The majority of the objective metrics were analyzed by SA level (overall ($SA_O$), perception ($SA_1$), comprehension ($SA_2$), and projection ($SA_3$)), decision difficulty (overall, easy, and hard), timing with respect to an SA probe question (15 seconds before asking, while being asked, or during response to an SA probe question), or per participant. Non-parametric statistical methods, including Mann-Whitney-Wilcoxon tests with one
degree of freedom (DOF = 1) and Spearman Correlations, were calculated due to a lack of normality. The correlations were with respect to SA Probe Accuracy and Selection Success Rate. The Collective evaluation data was reanalyzed using the same methods. Secondary research questions’ hypotheses, associated metrics, results, and discussion are presented for each research question in Sections 5–8.

5 R1: SYSTEM DESIGN ELEMENT INFLUENCE ON HUMAN OPERATOR

Understanding how the combined model and visualization influenced the operator, R1, is necessary to determine if the transparency embedded into the system design aided operators with different capabilities. The associated objective dependent variables were (1) the operator’s ability to influence the collective robotic system to choose the highest-valued target, (2) SA, (3) visualization clutter, (4) the operator’s spatial reasoning capability, and (5) the operator’s working memory capacity. The specific direct and indirect transparency factors related to R1 are identified in Figure 3. The relationship between the variables and the corresponding hypotheses, as well as the direct and indirect transparency factors, are identified in Table 1. Additional relationships (not identified in Figure 1) between the variable and transparency factors are identified due to correlation analyses.

The hypotheses in this section and the subsequent result Sections 6–8 are phrased using the $M_2$ model and Collective visualization, because each respective system design element provided the best transparency in the prior evaluations [11, 33]. It was hypothesized ($H_1$) that operators using $M_2$ and the Collective visualization will experience significantly higher SA and lower workload. SA represents an operator’s ability to perceive and comprehend information to project future actions [14]. Usability influences the perception of information [34] and will impact workload, which is the amount of stress an operator experiences to accomplish a task in a particular time duration [44]. It was hypothesized ($H_2$) that operators with different individual capabilities will not perform significantly different using $M_2$ and the Collective visualization. Ideal system design elements will enable operators with different capabilities to perceive, comprehend, and influence collectives relatively the same. The operator’s attitude and sentiments towards a system is dependent on system usability and provides information related to the system’s design [26]. Good designs promote higher operator satisfaction. It was hypothesized ($H_3$) that operators using $M_2$ and the Collective visualization will experience significantly less frustration (i.e., higher satisfaction).

Fig. 3. $R_1$ concept map of the assessed direct and indirect transparency factors.
Table 1. Interactions that Influence the Human Operator’s Objective (Obj) and Subjective (Subj) Variables (Vars), Relative to the Hypotheses (H), as Well as the Direct and Indirect Transparency Factors (See Figure 3)

| Transparency Factors | Direct |
|----------------------|--------|
|                      | Explainability | Observable | Performance | Understanding | Usability | Capability | Effectiveness | Predictability | Reliability | SA | Satisfaction | Timing | Workload |
|                      | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

| Obj Vars      | H     |
|---------------|-------|
| Target Value  | $H_1$ |
| SA Probe Accuracy | $H_1$ |
| Global Clutter | $H_1$ |
| Mental Rotations Assessment | $H_2$ |
| Working Memory Capacity | $H_2$ |

| Subj Vars                                      |
|-----------------------------------------------|
| Weekly Hours on a Desktop or Laptop           | $H_2$ |
| NASA-TLX                                      | $H_1,H_3$ |
| Post-Experiment                               | $H_2$ |

5.1 Metrics and Results

Assessing variables, such as the selected target’s value, is necessary to determine whether operators perceived the target value correctly and influenced the collectives positively. The human-collective team’s objective was to select the highest-valued target for each decision from a range of target values (67 to 100). IA interface operators using the $M_2$ model chose higher-valued targets compared to $M_3$, regardless of the decision difficulty, while Collective interface operators using $M_3$ chose higher-valued targets for overall and hard decisions. The target value median, min, max, and the Mann-Whitney-Wilcoxon significant effects between models for each model and visualization combination are shown in Figure 4. IA interface operators had significantly different selected target values between models for overall and easy decisions, while no differences were found for the Collective interface operators. Between visualizations Mann-Whitney-Wilcoxon tests identified moderate significant effects using $M_3$ for overall decisions ($N = 672, U = 63,946, p < 0.01$) and highly significant effects for hard decisions ($N = 276, U = 12,058, p < 0.001$). Collective interface operators chose higher-valued targets compared to the IA interface operators.

The SA-dependent variable was SA probe accuracy, the percentage of correctly answered SA probe questions [10]. Each question corresponded to the three SA levels: perception, comprehension, and projection [14]. The five $SA_1$ questions determined the operator’s ability to perceive collective and target information, such as “What collectives are investigating Target 3?” The operator’s comprehension of information was determined by four $SA_2$ questions, such as “Which Collective has achieved a majority support for Target 7?” Three $SA_3$ questions were related to the operator estimating the collectives’ future state, such as “Will support for Target 1 decrease?” The overall SA value, $SA_O$, was the percent of correctly answered SA probes. Operators from both evaluations using $M_2$, when compared to $M_3$, had higher $SA_3$, while the IA interface operators had
higher $SA_2$, and the Collective interface operators had higher $SA_O$. The SA probe accuracy for each model and visualization combination are shown in Figure 5. Significant differences between models were found for IA interface operators answering $SA_1$ probe questions and for Collective interface operators answering $SA_3$ probe questions. Between visualizations Mann-Whitney-Wilcoxon tests ($N = 56$) identified highly significant effects using $M_2$ for $SA_O$ ($U = 702, p < 0.001$) and $SA_1$ ($U = 714.5, p < 0.001$); and moderately significant effects for $SA_2$ ($U = 572.5, p < 0.01$) and $SA_3$ ($U = 554, p < 0.01$). Highly significant effects between visualizations were found using $M_3$ for $SA_O$ ($U = 657.5, p < 0.001$), $SA_2$ ($U = 648, p < 0.001$), and $SA_3$ ($U = 645.5, p < 0.001$). A moderately significant effect between visualizations was found using $M_3$ for $SA_1$ ($U = 564, p < 0.01$). Collective interface operators had higher SA probe accuracy in general.

Global clutter percentages were analyzed for each SA probe question. Clutter is the area occupied by objects on a display, relative to the display’s total area [44]. Clutter becomes an issue when presenting too much information in close proximity, which requires a longer search time [44] and negatively influences the operator’s ability to perform a task. Area coverage was calculated as the number of pixels an item covered on the computer visualization. One meter for the IA visualization was approximately 1.97 pixels per meter. The Collective visualization computer display size was unknown; therefore, global clutter percentage calculations used the IA visualization corresponding item and computer display dimensions (2,073,600 pixels$^2$). The global clutter percentage variable was the percentage of computer screen area obstructed by all displayed objects,
using Equation (1):

\[
\text{GlobalClutter}(\%) = \left( \frac{\text{ICA} + \text{GHA} + \text{GHTA} + \text{GTA} + \text{GAICE} + \text{GTIW} + \text{GCIW}}{2,073,600} \right) \cdot 100, \tag{1}
\]

where ICA represents the static interface components areas (493,414 pixels\(^2\)). GHA represents the area covered by Collectives I-IV (9,856 pixels\(^2\)), which were visible throughout the trial. The area corresponding to highlighted targets (2,350 pixels\(^2\) per highlighted target), which have outlines and are in range of the selected collective, are represented as GHTA. Remaining targets that are not highlighted (1,720 pixels\(^2\) per target) are represented as GTA. GAICE is the area consumed by the 800 individual collective entities (51,200 pixels\(^2\)), only considered for the IA visualization. The area corresponding to the number of target information windows (32,922 pixels\(^2\) per target information window) is represented as GTIW, and the corresponding collective information windows is represented as GCIW (25,740 pixels\(^2\) per collective information window). The clutter associated with the background map was not considered in the global clutter calculation for two reasons. First, the underlying map is identical for both visualizations, but differed slightly due to the computer screen size, making a between-evaluation assessment unattainable. Second, the operators did not depend on the underlying map to complete the sequential best-of-n task. The map provided an ecologically valid background, representative of the task in a real-world dynamic environment. Future evaluations must consider how the background may influence clutter, especially if the map becomes dynamic, or the operator can zoom in and out on particular areas.

IA interface operators using \(M_2\) had lower global clutter percentages compared when using \(M_3\). Collective interface operators had lower global clutter percentages using \(M_2\). \(SA_3\) at all timings and \(SA_1\), while being asked an SA probe question were lower when Collective interface operators used \(M_3\). The global clutter percentage for each model and visualization combination are shown in Figure 6. Significant differences between models were found for IA interface operators at all timings for \(SA_O\), \(SA_1\), and \(SA_2\) probe questions, while a significant difference between models was identified for Collective interface operators during response to \(SA_2\) probe questions. Between-visualizations Mann-Whitney-Wilcoxon tests identified significant differences using \(M_5\). A highly significant effect between visualizations was found responding to an SA probe question for \(SA_O\) (\(N = 670, U = 64,442, p < 0.001\)). Moderate significant effects between visualizations were found for \(SA_O\) 15 seconds before asking (\(U = 64,188, p < 0.01\)) and while being asked an SA probe question (\(U = 63,728, p < 0.01\)). Significant effects between visualizations were found 15 seconds before asking for \(SA_1\) (\(N = 294, U = 12,487, p = 0.02\)) and \(SA_3\) (\(N = 152, U = 3,445.5, p = 0.03\)); while being asked for \(SA_1\) \(U = 12,301, p = 0.03\) and \(SA_3\) \(U = 3452, p = 0.05\); and during response to an SA probe question for \(SA_1\) \(U = 12,216, p = 0.04\). Correlations between the global clutter percentage and SA probe accuracy were found using the Collective visualization 15 seconds before asking an SA probe question. The Spearman correlation analysis found a moderate correlation using \(M_3\) for \(SA_3\) \((r = 0.45, p < 0.001)\) and weak correlations using \(M_2\) for \(SA_1\) \((r = 0.16, p = 0.05)\) and for \(SA_O\) using \(M_3\) \((r = 0.2, p < 0.001)\). The IA visualization had lower global clutter percentages compared to the Collective visualization. Collective interface operators using \(M_3\), however, had lower global clutter while being asked and during response to an \(SA_1\) probe question.

There were no significant effects between visualizations for the Mental Rotations Assessment [41]. Correlations between the Mental Rotations Assessment and SA probe accuracy existed for the IA visualization. The Spearman correlation analysis revealed weak correlations with \(M_2\) for \(SA_O\) \((r = 0.17, p < 0.01)\), \(SA_1\) \((r = 0.18, p = 0.03)\), and \(SA_2\) \((r = 0.27, p < 0.01)\). Weak correlations were revealed with \(M_3\) for \(SA_O\) \((r = 0.15, p < 0.01)\), \(SA_1\) \((r = 0.19, p = 0.03)\), and \(SA_2\) \((r = 0.18, p = 0.05)\). A moderate correlation existed between Working Memory Capacity and SA probe accuracy for the IA visualization using \(M_2\) for \(SA_3\) \((r = 0.45, p < 0.001)\). Weak correlations existed with \(M_2\)
for $SA_0$ ($r = 0.23, p < 0.001$) and $SA_1$ ($r = 0.17, p = 0.04$), and using $M_3$ for $SA_O$ ($r = 0.14, p = 0.01$). The Mann-Whitney-Wilcoxon tests found no significant effects between visualizations for the weekly hours spent using a desktop or laptop. Weak correlations were found between weekly hours using a desktop or laptop and SA probe accuracy using $M_2$ for the IA visualization for $SA_O$ ($r = 0.12, p = 0.04$) and $SA_1$ ($r = 0.21, p = 0.01$), and for the Collective visualization for $SA_2$ ($r = 0.21, p = 0.02$).
The NASA Task Load Index (NASA-TLX) assessed six workload subscales and the overall workload [23]. IA interface operators using $M_2$ had lower physical demand and effort compared to $M_3$, while Collective interface operators had lower physical demand, effort, and frustration when using $M_2$. The NASA-TLX for each model and visualization combination are shown in Figure 7. IA interface operators had significantly different rankings between models for physical demand and frustration, while mental demand was significantly different between models for Collective interface operators. Between visualizations Mann-Whitney-Wilcoxon tests ($N = 56$) identified a significant effect using $M_2$ for mental demand ($U = 515, p = 0.04$) and a highly significant effect for performance ($U = 159.5, p < 0.001$). Significant effects were found between visualizations using $M_3$ for overall workload ($U = 266.5, p = 0.04$), performance ($U = 242.5, p = 0.01$), and frustration ($U = 511, p = 0.05$), as well as a highly significant effect for physical demand ($U = 208, p < 0.001$). The Collective visualization imposed a lower overall workload, had lower physical and temporal demands, and caused less frustration compared to the IA visualization.

The post-experiment questionnaire assessed the collective’s request responsiveness, the operators’ ability to choose the highest-valued target, and their understanding of collective behavior, from best (1) to worst (2 for the IA evaluation and 3 for the Collective evaluation). The best collective responsiveness, operator ability, and understanding occurred when IA interface operators used $M_2$ versus $M_3$. Collective interface operators ranked the collective’s responsiveness highest using $M_3$ and the operator ability and understanding using $M_2$. The post-experiment for each combined model and visualization are shown in Figure 8. System responsiveness, operator ability, and understanding were significantly different between models for IA interface operators, while system responsiveness and operator understanding were significantly different for Collective interface operators.

A summary of $R_1$’s results that show the hypotheses with associated significant results is provided in Table 2. This summary table is intended to facilitate the discussion.

5.2 Discussion

Relationships to the transparency factors provided in Table 1 are emphasized using italics. The analysis of how the combined model and visualization influenced operators with different individual capabilities suggests that the $M_2$ model promoted transparency as effectively as $M_3$, while the Collective visualization promoted better transparency compared to the IA visualization. $H_1$, which hypothesized that operators using $M_2$ and the Collective visualization will experience significantly higher SA and lower workload, was not supported. SA performance (i.e., accuracy) varied across the SA levels, depending on the model, and workload varied across the workload subscales,
Table 2. A Synopsis of $R_1$’s Hypotheses Associated with Significant Results

| Variable | Sub-Variable | Within Model | Between Visualization | Correlation |
|----------|--------------|--------------|-----------------------|-------------|
|          |              | IA Coll.     | $M_2$ $M_3$         |             |
|          |              | $M_2$ $M_3$ | $M_2$ $M_3$         |             |
| Target Value | Overall | $H_1$ | $H_1$ |         |             |
|          | Hard        | $H_1$ | $H_1$ |         |             |
|          | $SA_O$      | $H_1$ | $H_1$ |         |             |
|          | $SA_1$     | $H_1$ | $H_1$ |         |             |
|          | $SA_2$     | $H_1$ | $H_1$ |         |             |
|          | $SA_3$     | $H_1$ | $H_1$ |         |             |
| SA Probe Accuracy | $SA_O$ | $H_1$ | $H_1$ | $H_1$ |             |
|          | $SA_1$     | $H_1$ | $H_1$ | $H_1$ |             |
|          | $SA_2$     | $H_1$ | $H_1$ | $H_1$ |             |
|          | $SA_3$     | $H_1$ | $H_1$ | $H_1$ |             |
| Global Clutter | $SA_O$ | $H_1$ | $H_1$ |         |             |
|          | $SA_1$     | $H_1$ | $H_1$ | $H_1$ |             |
|          | $SA_2$     | $H_1$ | $H_1$ | $H_1$ |             |
|          | $SA_3$     | $H_1$ | $H_1$ | $H_1$ |             |
| Mental Rotations Assessment | $SA_O$ | $H_2$ | $H_2$ |         |             |
|          | $SA_1$     | $H_2$ | $H_2$ |         |             |
|          | $SA_2$     | $H_2$ | $H_2$ |         |             |
| Working Memory Capacity | $SA_O$ | $H_2$ | $H_2$ |         |             |
|          | $SA_1$     | $H_2$ | $H_2$ |         |             |
|          | $SA_3$     | $H_2$ | $H_2$ |         |             |
| Weekly Hours on Desktop or Laptop | $SA_O$ | $H_2$ | $H_2$ |         |             |
|          | $SA_1$     | $H_2$ | $H_2$ |         |             |
|          | $SA_2$     | $H_2$ | $H_2$ |         |             |
| NASA-TLX | Overall | $H_1$ |         |         |             |
|          | Mental      | $H_1$ |         | $H_1$ |             |
|          | Physical    | $H_1$ |         | $H_1$ |             |
|          | Performance | $H_1$ |         | $H_1$ |             |
|          | Frustration | $H_1$, |         | $H_1$, |             |

The timings are all timings (All), 15 seconds Before asking (B), While being asked (W), and During response (D) to an SA probe question.

depending on the model and visualization combination. $M_2$ was effective at enabling operators to more accurately predict future collective robotic system behaviors, while $M_3$ enabled better observability of the collectives’ behaviors. Better predictability may have occurred, because $M_2$ aligned with the operators’ expectations: that the model was designed to choose the highest-valued target. Predictability of future collective states may have also improved due to the visualization. Favoring entities in the IA visualization created streamlines between hubs and targets, which may have directed the operator’s attention to particular targets. $M_3$ may have promoted better perception of the collectives’ behaviors, because achieving the task required the operator to direct the behaviors. Operator workload was alleviated by $M_2$ by requiring fewer operator capabilities, such as
physical demand and effort, as well as promoting higher performance, which was expected, since operator influence was not required to make decisions. \( M_3 \) alleviated operator workload by also requiring fewer operator capabilities, such as mental demands, and improving satisfaction (i.e., less frustration) by mitigating temporal (i.e., timing) demands. More operator control, such as making decisions quickly or more slowly, may have contributed to these lower workload subscales.

Transparency embedded into the Collective visualization partially supported \( H_1 \), because it promoted higher SA performance via the color-coded icons and outlines, state information identified on the collective icon, information provided in the pop-up windows, and feedback provided in the Collective Assignments and System Messages areas. Collective interface operators encountered more clutter due to the long duration of time the target information windows were visible. The increased clutter had both positive and negative transparency implications. Clutter, from a usability perspective, is not ideal if operators are unable to perform their tasks effectively. The Collective interface operators with higher clutter answered more SA probe questions accurately, which suggests that the operators were not hindered by the clutter and performed better. The dependence on the visible target information windows may have been attributed to the type of SA probe question. Thirteen of 24 SA probe questions relied on information provided in the target information windows. An example question, such as "What collectives are investigating Target 3?", required using the target information window if Target 3 was in range of multiple collectives. The operator was able to identify which collectives were within range of a target by left-clicking on the respective collective; however, target information windows were needed to see the numeric collective support values from the collectives. Experimental design modifications can ensure a more evenly distributed SA probe questions that rely on other information, such as the icons, system messages, or collective assignments versus information windows. Target icon design modifications that indicate which collectives support a particular target may improve explainability, reliability, and increase the reliance on the target icon instead of the information window.

The Collective visualization partially supported \( H_1 \) by requiring fewer operator capabilities, such as physical demands, and improving satisfaction (i.e., less frustration) by mitigating temporal (i.e., timing) demands. Not visualizing entities may have reduced operator stress, because the rate of a collective’s state change was not easily perceived. The need or desire to influence collective behaviors may not have been as apparent, which attributed to lower physical demand and frustration. Higher operator mental demand when using \( M_2 \) and the Collective visualization may have occurred if collective behaviors, or state changes, were not observable and required more interactions to deduce what was happening, such as accessing information windows.

\( H_2 \), which hypothesized that operators with different individual capabilities did not perform significantly different using \( M_2 \) and the Collective visualization, was partially supported. Individuals with different capabilities, spatial reasoning, and working memory capacity, performed relatively the same. Operators who had more computer knowledge, however, better understood the collective behaviors. This finding was anticipated. Future investigations will identify what particular aspects of computer knowledge resulted in better understanding.

Collective operators using \( M_2 \) were more satisfied, which supported \( H_3 \). Dissatisfaction transpires when the system is not transparent and prohibits the operator from understanding what is happening or the interface appears visually noisy due to clutter [30]. An autonomous model, one with decision-making capabilities, and an abstract collective visualization may mitigate dissatisfaction. More metrics, such as the Questionnaire for User Interface Satisfaction [9], are needed to assess how the transparency embedded in the combined models and visualizations influence operator satisfaction.

The transparency embedded in the Collective visualization supported operators with individual differences better than the IA visualization. \( M_2 \), however, did not support all operators. More
computer experience, for example, aided operator SA performance. Mitigating the need for operators to have particular capabilities is desired to design effective human-collective systems. Higher SA performance varied between the models, which suggests system design changes must be considered to improve the perception, comprehension, and projection of future collective behaviors when using \( M_2 \). Usability considerations need to identify the ideal amount of operator influence in the decision-making process to alleviate workload (e.g., mental demand) and promote better SA.

6 \( R_2 \): SYSTEM DESIGN ELEMENT PROMOTION OF OPERATOR COMPREHENSION

\( R_2 \)'s exploration of the explainability direct transparency factor focused on determining whether the transparency embedded in the combined model and visualization promoted operator comprehension. Perception and comprehension of the information are necessary to inform future operator actions. The associated objective dependent variables were (1) SA, (2) collective and target left- or right-clicks, (3) collective and target observations, (4) interventions, (5) the percentage of times the highest-valued target was abandoned, and (6) whether the information window was open when a target was abandoned. The specific direct and indirect transparency factors related to \( R_2 \) are identified in Figure 9. The relationship between the variables and the corresponding hypotheses, as well as the direct and indirect transparency factors, are identified in Table 3. Relationships between the variable and the direct or indirect transparency factors that are not shown in Figure 1 were identified after conducting correlation analyses.

Models designed to aid operators to fulfill a best-of-\( n \) decision-making task can help mitigate workload by reducing repetitive interactions, ensuring task progress if an operator is distracted, and allowing more time to establish situational awareness and understanding. Display principles, associated with perceptual operations, mental models, as well as human attention and memory [44], may improve understanding by providing legible, clear, concise, organized, easily accessible, and consistent information. Providing information, such as the collective robotic system state, on the collective’s hub, rather than using all of the individual collective entities, is more clear, concise, organized, and consistent. It was hypothesized (\( H_4 \)) that operators will have a better understanding of the \( M_2 \) model and information provided by the Collective visualization. Appropriate expectations of the model’s capabilities and contributions towards a goal, as well as providing information redundantly via icons, colors, messages, and the collective and target information windows can aid operator comprehension and justify their future actions. It was hypothesized (\( H_5 \)) that operators.

---

**Fig. 9.** \( R_2 \) concept map of the assessed direct and indirect transparency factors.
Table 3. Interactions That Influence the Human Operator’s Objective and Subjective Variables, Relative to the Hypotheses, as Well as the Direct and Indirect Transparency Factors (See Figure 9)

| Transparency Factors |
|-----------------------|
| **Obj Vars** | **H** | Directable | Explainability | Information | Observable | Performance | Understanding | Usability | Capability | Control | Effectiveness | Justification | Predictability | Reliability | SA | Timing |
| SA Probe Accuracy | $H_4$ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Collective Left-clicks by SA Lvl | $H_5$ | ✓ | ✓ | ✓ | ✓ |
| Target Right-clicks by SA Lvl | $H_5$ | ✓ | ✓ | ✓ | ✓ |
| Collective Observations | $H_5$ | ✓ | ✓ | ✓ | ✓ |
| Target Observations | $H_4$ | ✓ | ✓ | ✓ | ✓ |
| Collective Right-clicks per Dec. | $H_5$ | ✓ | ✓ | ✓ | ✓ |
| Target Right-clicks per Dec. | $H_5$ | ✓ | ✓ | ✓ | ✓ |
| Interventions | $H_4$ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Highest Value | $H_4$, $H_5$ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Target Abandon | $H_5$ | ✓ | ✓ | ✓ | ✓ |
| Abandon Target Information Window Open | $H_5$ | ✓ | ✓ | ✓ | ✓ |

| Subj Vars |
|-----------------------|
| **Post-trial Performance and Understanding** | $H_4$ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Post-experiment** | $H_4$ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

using $M_2$ and the Collective visualization were able to accurately justify their actions. An ideal system will enable operators to perceive and comprehend information that is explainable, which will support effective human-collective interactions.

### 6.1 Metrics and Results

The operator had access to supplementary information that was not displayed continually, such as different colored target borders that identified which targets were in range and had been abandoned or information windows that provided collective robotic system state and target support information, to aid comprehension ($SA_2$) of collective behavior and inform actions. The results of SA probe accuracy, which is the percentage of correctly answered SA probes questions used to assess the operator’s SA, identified that IA and Collective interface operators using the $M_2$ model...
Collective left-clicks identified all targets that were within range of a collective and was the first click required to issue a command. $M_2$ in general had fewer collective left-clicks compared to $M_3$. Collective interface operators using $M_3$ while being asked an SA probe question had fewer collective left-clicks for $SA_O$. The number of collective left-clicks median, min, max, and the Mann-Whitney-Wilcoxon significant effects between models are presented in Figure 10. IA interface operators had significantly different collective left-clicks between models for $SA_O$, $SA_1$, and $SA_2$ at all timings, as well as $SA_3$ 15 seconds before asking and during response to an SA probe question. Significantly different collective left-clicks between models were identified 15 seconds before asking an SA probe question for $SA_O$, $SA_2$, and $SA_3$, while being asked an SA probe question for $SA_O$ and $SA_1$, and during response to an SA probe question for $SA_O$, $SA_1$, and $SA_2$.

Additional between-visualizations Mann-Whitney-Wilcoxon tests identified highly significant effects when using $M_2$ 15 seconds before asking an SA probe question for $SA_O$ ($N = 664$, $U = 64,213$, $p < 0.001$), a moderate significant effect for $SA_1$ ($N = 290$, $U = 12,534$, $p < 0.01$), and a significant effect for $SA_2$ ($N = 223$, $U = 7,210.5$, $p = 0.04$). Highly significant effects between visualizations when using $M_2$ while being asked an SA probe question were found for $SA_O$ ($U = 67,670$, $p < 0.001$), and $SA_2$ ($U = 8,317$, $p < 0.001$), as were moderately significant effects for $SA_1$ ($U = 12,043$, $p < 0.01$), and $SA_3$ ($N = 151$, $U = 3,472$, $p < 0.01$). A highly significant effect between visualizations when
using $M_2$ during response to an SA probe question was found for $SA_O$ ($U = 64,710, p < 0.001$), a moderate significant effect for $SA_1$ ($U = 12,414, p < 0.01$), and a significant effect for $SA_3$ ($U = 3,489, p = 0.01$). A significant effect between visualizations when using $M_3$ 15 seconds before asking an SA probe question was found for $SA_O$ ($N = 665, U = 60,696, p = 0.03$). Highly significant effects between visualizations when using $M_3$ while being asked an SA probe question were found for $SA_O$ ($U = 64,376, p < 0.001$), $SA_1$ ($N = 251, U = 9,959.5, p < 0.001$), as well as a moderate significant effect for $SA_3$ ($N = 162, U = 4,114, p < 0.01$). Correlations between the collective left-clicks and SA probe accuracy were only revealed when using $M_3$. The Spearman correlation analysis revealed weak correlations for the IA visualization for $SA_3$ 15 seconds before asking ($r = -0.26, p = 0.02$) and while being asked an SA probe question ($r = -0.33, p < 0.01$). Weak correlations were also revealed for the Collective visualization while being asked an SA probe question for $SA_O$ ($r = 0.13, p = 0.02$) and $SA_1$ ($r = 0.22, p = 0.02$). The IA visualization had fewer collective left-clicks in general compared to the Collective visualization. Collective interface operators who used $M_2$ during response to an SA probe question had fewer left-clicks for $SA_O$, and when using $M_3$ for all SA levels.

Target right-clicks allowed operators to access target information windows that provided each collective’s support percentage for a respective target, which may have been used to justify issuing commands. $M_2$ had fewer target right-clicks for both visualizations. Collective interface operators who used $M_3$ while being asked an SA probe question had fewer target right-clicks for $SA_3$. The number of target right-clicks for each combined model and visualization are shown in Figure 11. IA interface operators had significantly different target right-clicks between models 15 seconds before asking an SA probe question for $SA_O$, $SA_2$, and $SA_3$, as well as during response to an SA
probe question for $SA_O$ and $SA_3$. Significantly different target right-clicks between models were found 15 seconds before asking an SA probe question for $SA_O$ and during response to an SA probe question for $SA_O$ and $SA_1$. No significant effects between visualizations were found. The Collective interface operators using $M_2$ had fewer target right-clicks for all SA levels, 15 seconds before asking and during response to an SA probe question compared to the IA interface operators. Fewer target right-clicks, 15 seconds before asking and while being asked an SA probe question, occurred when IA interface operators used $M_3$ compared to the Collective interface operators. The Spearman correlation analysis found weak correlations between target right-clicks and SA probe accuracy for the IA interface operators using $M_2$ 15 seconds before asking an SA probe question for $SA_O$ ($r = 0.17, p < 0.01$) and $SA_2$ ($r = 0.37, p < 0.001$). Weak correlations were found for the IA interface operators using $M_3$ 15 seconds before asking an SA probe question for $SA_O$ ($r = 0.11, p = 0.04$), $SA_1$ ($r = 0.2, p = 0.02$), and for the Collective interface operators for $SA_1$ 15 seconds before asking ($r = -0.24, p = 0.01$) and while being asked an SA probe question ($r = -0.21, p = 0.03$).

Collective observations were collective left-clicks that identified targets within range of a collective (i.e., white borders indicated that the individual entities were investigating the target, while yellow indicated no investigation) and targets that were abandoned (i.e., red borders). IA interface operators using $M_3$ had fewer collective observations compared to $M_2$, while Collective interface operators had fewer collective observations when using $M_2$. The collective observations for both models at all decision difficulties had median = 100, min = 0, and max = 100. IA interface operators had significantly different collective observations between models for all decision difficulties, while Collective interface operators had significantly different collective observations between models for easy decisions. Between-visualizations Mann-Whitney-Wilcoxon tests identified a moderate significant effect using $M_2$ for overall decisions ($N = 672, U = 61,152, p < 0.01$) and a significant effect for easy decisions ($N = 374, U = 19,008, p = 0.05$). Highly significant effects between visualizations were found using $M_3$ for overall ($U = 73,920, p < 0.001$), easy ($N = 396, U = 25,587, p < 0.001$), and hard decisions ($N = 276, U = 12,520, p < 0.001$). The IA visualization had fewer collective observations compared to the Collective visualization.

Target observations were target left-clicks not associated with issuing a command. $M_2$ and the Collective visualization had fewer target observations, regardless of decision difficulty. The target observations for both models at all decision difficulties had min = 0 and max = 100 with a median = 100 for IA interface operators and median = 0 for Collective interface operators. IA interface operators had significantly different target observations between models for all decision difficulties. Between-visualizations Mann-Whitney-Wilcoxon tests identified highly significant effects when using $M_2$ for overall ($N = 672, U = 35,280, p < 0.001$), easy ($N = 374, U = 10,886, p < 0.001$), and hard decisions ($N = 298, U = 6,910, p < 0.001$). Highly significant effects between visualizations when using $M_3$ were also found for overall ($U = 41,664, p < 0.001$), easy ($N = 396, U = 15,053, p < 0.001$), and hard decisions ($N = 276, U = 6,615, p < 0.001$).

Collective right-clicks allowed the operator to access information windows, which provided the number of individual entities in each state. Operators may have used the information to justify issuing commands. The number of collective right-clicks was only assessed for the IA evaluation, because the Collective evaluation did not record which collective window was opened or closed. $M_3$ had fewer collective right-clicks compared to $M_2$, regardless of decision difficulty. The collective right-clicks for each combined model and visualization are shown in Figure 12(a). Significantly different collective right-clicks between models were found for overall and hard decisions.

Target right-clicks allowed the operator to access target information windows, which provided the percentage of support each collective had for a respective target. The target support information may have also been used to justify issuing commands, such as increasing or decreasing
Fig. 12. Right-clicks per decision by decision difficulty.

Table 4. Interventions (Abandoned Targets with 10% Support) per Participant Descriptive Statistics

| Model  | Mean (SD) | Median (Min/Max) |
|--------|-----------|------------------|
| IA     |           |                  |
| $M_2$  | 1.5 (2.03)| 0.5 (0/7)        |
| $M_3$  | 3.75 (4.27)| 3 (0/17)        |
| Collective |      |                  |
| $M_2$  | 2.21 (1.99)| 1.5 (0/7)       |
| $M_3$  | 5 (5.11)  | 3.5 (0/18)       |

support from particular collectives. IA interface operators using $M_2$ had fewer target right-clicks, while Collective interface operators had fewer target right-clicks using $M_3$. The target right-clicks for each combined model and visualization are shown in Figure 12(b). IA interface operators had significantly different target right-clicks between models for easy decisions, while no differences were found for the Collective interface operators. The Collective visualization had fewer target right-clicks compared to the IA visualization; however, no significant effects between visualizations were found.

Interventions occurred when the operator abandoned a target with greater than 10% collective support. Abandoning low-value targets was a desired intervention. Interventions were assessed per participant, due to the inability to associate an intervention to a decision, and the descriptive statistics are shown in Table 4 [11]. $M_2$ and the IA visualization had fewer interventions. The Mann-Whitney-Wilcoxon tests found a significant effect between models for the IA visualization ($N = 56, U = 270.5, p = 0.04$). No significant effects between visualizations were found.

The abandon command discontinued a collective’s investigation of a particular target. Ideally, lower-valued targets were abandoned, since the objective was to aid each collective in selecting and moving to the highest-valued target. Operators using $M_3$ abandoned the highest-valued target less frequently compared to $M_2$. The highest-valued target abandoned for each combined model and visualization is shown in Figure 13(a). IA interface operators had significantly different highest-valued target abandoned percentages between models for easy decisions, while Collective interface operators had significant differences between models for overall decisions. Operators using the IA visualization abandoned the highest-valued target less frequently compared to those using the Collective visualization; however, no significant effects were found between the visualizations.

The percentage of times an abandoned target information window was open per participant was evaluated. The operator may have used the support information to justify abandoning a target. Operators using $M_3$ had fewer abandoned target information windows open compared to $M_2$. The
abandoned target information window open for each combined model and visualization are shown in Figure 13(b), but no significant effects between models were found. Between-visualizations Mann-Whitney-Wilcoxon tests identified significant effects when using $M_3$ for overall ($N = 49$, $U = 414.5$, $p = 0.02$) and easy decisions ($N = 45$, $U = 352$, $p = 0.02$). Fewer abandoned target information windows were open when using the IA visualization compared to the Collective visualization.

The post-trial questionnaire assessed the operators’ understanding of collective behavior and their ability to choose the best target, never (1) to always (7). Performance and understanding were higher for Collective interface operators using $M_3$. The post-trial performance and understanding for each combined model and visualization are shown in Figure 14. IA interface operators ranked understanding significantly different between models. Between-visualizations Mann-Whitney-Wilcoxon tests found a significant effect for understanding using $M_2$ ($N = 56$, $U = 513$, $p = 0.04$).

The post-experiment questionnaire assessed the collective’s request responsiveness, the participants’ ability to choose the highest-valued target, and their understanding of the collective behavior. IA interface operators using $M_2$ had the best collective responsiveness, operator ability, and understanding versus $M_3$. Collective interface operators ranked the collective’s responsiveness highest using $M_3$, while operator ability and understanding were highest using $M_2$. Statistical test details were provided in the Metrics and Results Section 5.1.

A summary of $R^2$’s results that shows the hypotheses with associated significant results is provided in Table 5. This summary table is intended to facilitate the discussion.
Table 5. A Synopsis of $R_2$'s Hypotheses Associated with Significant Results

| Variable                        | Sub-Variable | Within Model | Between Model | Correlation |
|---------------------------------|--------------|--------------|---------------|-------------|
|                                 |              | IA | Coll. | $M_2$ | $M_3$ | IA | Coll. | $M_2$ | $M_3$ |
| SA Probe Accuracy               | $SA_0$       | $H_5$ | -All | $H_5$ | $-All$ | $H_5$ | -B,W | $H_5$ | -W  |
|                                 | $SA_1$       | $H_5$ | -All | $H_5$ | -W,D | $H_5$ | -W  | $H_5$ | -W  |
|                                 | $SA_2$       | $H_5$ | -All | $H_5$ | -B,D | $H_5$ | -B  | $H_5$ | -B  |
|                                 | $SA_3$       | $H_5$ | -B,D | $H_5$ | -B   | $H_5$ | -W  | $H_5$ | -W  |
| Collective Left-clicks          | $SA_0$       | $H_5$ | -All | $H_5$ | -B,D | $H_5$ | -B  | $H_5$ | -B  |
|                                 | $SA_1$       | $H_5$ | -D   | $H_5$ | -D   | $H_5$ | -D  | $H_5$ | -D  |
|                                 | $SA_2$       | $H_5$ | -B   | $H_5$ | -B   | $H_5$ | -B  | $H_5$ | -B  |
|                                 | $SA_3$       | $H_5$ | -B,D | $H_5$ | -B,D | $H_5$ | -B,D | $H_5$ | -B,D |
| Target Right-clicks by SA Level | Overall       | $H_5$ |       | $H_5$ |       | $H_5$ |       | $H_5$ |       |
|                                 | Easy          | $H_5$ |       | $H_5$ |       | $H_5$ |       | $H_5$ |       |
|                                 | Hard          | $H_5$ |       | $H_5$ |       | $H_5$ |       | $H_5$ |       |
| Target Observations             | Overall       | $H_4$ |       | $H_4$ |       | $H_4$ |       | $H_4$ |       |
|                                 | Easy          | $H_4$ |       | $H_4$ |       | $H_4$ |       | $H_4$ |       |
|                                 | Hard          | $H_4$ |       | $H_4$ |       | $H_4$ |       | $H_4$ |       |
| Collective Right-clicks per Decision | Overall       | $H_5$ |       | $H_5$ |       | $H_5$ |       | $H_5$ |       |
|                                 | Easy          | $H_5$ |       | $H_5$ |       | $H_5$ |       | $H_5$ |       |
| Interventions                   | Highest Value | $H_4$ |       |       |       | $H_5$ |       |       |       |
|                                 | Target Abandoned | Overall | $H_4$ | $H_5$ | $H_4$ | $H_5$ |       |       |       |
|                                 |               | Easy   | $H_4$ | $H_5$ |       |       |       |       |       |
| Abandoned Target Info. Window Open | Overall       |       |       | $H_5$ |       | $H_5$ |       |       |       |
|                                 | Easy          |       |       | $H_5$ |       | $H_5$ |       |       |       |
| Post-trial                      | Understand.   |       |       | $H_4$ |       | $H_4$ |       |       |       |
| Post-experiment                 | Understand.   |       |       | $H_4$ |       | $H_4$ |       |       |       |

The SA timings are all timings (All), 15 seconds Before asking (B), While being asked (W), and During response (D) to an SA probe question.
6.2 Discussion

The analysis of how the combined model and visualization promoted operator comprehension (i.e., the operator’s capability of understanding) suggests that the M$_3$ model promoted transparency more effectively than M$_2$, while both visualizations had their respective advantages and disadvantages. Operators using M$_2$ had fewer undesired interactions, such as target observations (i.e., extra clicks that did not contribute to the task) and interventions. Fewer undesired interactions may have occurred, because M$_2$ was designed to fulfill the best-of-$n$ decision-making task with or without operator influence, which effectively balanced control between the collective robotic systems and operator, whereas M$_3$ relied on operator influence (directability) to make a decision. More undesirable interactions, such as target observations, resulted in better task performance for operators when using M$_3$, which suggests that some interactions deemed undesirable for one model may be advantageous for another. Target observations may have occurred due to poor interface and visualization usability. Operators who issued commands (1) selected the desired command, (2) selected the desired collective and target, and (3) clicked on the commit button. Reissuing the same command required re-selecting the target and clicking the commit button. More target observations may have occurred if operators forgot to re-select the target when reissuing the same commands. Design improvements, such as leaving the target selected, may decrease target observations.

$H_4$, which hypothesized that operators will have a better understanding of M$_2$, was not supported, because operators using M$_2$ abandoned the highest-valued target more frequently. The operators may have been overloaded supervising the four collectives simultaneously, especially if they were distracted by the secondary task. The interface’s 10 (Hz) update rate (i.e., timing) may have negatively impacted the operator’s capability to understand what the collectives were doing and planned (e.g., predictability) to do. Introducing timing delays may afford operators more time to understanding the current situation; however, task completion will be prolonged, which is undesired in missions that require fast system responses. Providing predictive collective behaviors instead of timing delays may mitigate the time required for an operator to get back into-the-loop.

The highest-valued target was abandoned more frequently when using the Collective visualization. The target value may not have been observable enough (i.e., salient) to distinguish it from other potential targets, which did not support $H_4$. Future investigations will determine if the target value must use the entire collective hub icon area, similar to the IA visualization, to be more salient, and to establish what levels of obscurity are needed to ensure the target values are reliably distinguishable. Making distinctions clearer, such as using integers, to identify collectives versus targets, may improve visualization explainability and mitigate mistakes when operators confused the roman numeral identifiers with the integer identifiers. IA operators experienced this mistake frequently, which may have contributed to lowering their understanding. Ensuring that identifiers are unique and distinct will improve the effectiveness of the SA probe questions.

The use of target borders (collective observations), information windows (target right-clicks), and target value to determine if operators used this information to justify actions reliably. Collective interface operators using M$_2$ made better decisions with fewer collective observations and more target-right clicks. Understanding which collectives supported targets, by seeing numerical percentages, was more valuable compared to outlines indicating which targets were within a collective’s range. $H_5$, which hypothesized that operators using M$_2$ and the Collective visualization were able to justify actions accurately, was not supported. Collective interface operators who issued more collective left-clicks while being asked an SA probe question had better perception when using M$_3$. IA interface operators who issued more target right-clicks 15 seconds before asking an SA probe question had better comprehension when using M$_2$. The operators interactions were accurate and justified; however, the model and visualization combination did not support the
hypothesis. Collective left-clicks can improve perception of targets in range of a collective and are attributed with issued commands, which require perception, comprehension, and projection. Target right-clicks provide more information about collective support for a target, which may improve understanding.

Lower SA performance may have occurred if operators were in the middle of an interaction when the SA probe question was posed, while higher SA performance may have occurred because the operators anticipated when a SA probe question was going to be asked and took preventative actions. Operators using target information windows to verify that a target was abandoned may have been confused if the reported target support was greater than zero. There were instances during the trial when a few individual entities became lost, as the collective hub transitioned to a new location, and they did not move with the hub. The lost entities may have continued to explore an abandoned target, because they never received the abandon target message, which occurred inside of the hub. The operators, as a result, may have reissued additional abandon commands in an attempt to reduce the collective support to zero, although only one abandon command was needed. Strategies improving explainability, such as reporting zero percent support when an abandon command is issued and identifying how many individual entities have been lost, may mitigate erroneous repeated abandon command behavior and improve understanding. IA operators may have also experienced confusion if they saw individual entities still travelling to an abandoned target. Not displaying lost entities after a specific period of time, once a collective hub has moved to a new location, may reduce the number of reissued abandon commands. Further analysis using eye-tracking technology may provide more reliable metrics to determine operator comprehension by identifying exactly where an operator is focusing their attention.

The transparency embedded in combined M2 and the Collective visualization did not support the operator’s capability to understand the collectives’ behaviors the best. M3 provided better operator comprehension, because operators were more involved in the decision-making process. More interactions, even when undesirable, contributed to better understanding and task performance. Strategies to increase operator involvement, without requiring complete control over the decision-making process, using M2 must be considered to improve its effectiveness. Design improvements, such as increasing explainability by identifying how many individual entities became lost during a hub transition, can help mitigate abandoning the highest-valued target, which occurred frequently for Collective interface operators using M2. Understanding why particular interactions occurred for specific model and visualization combinations, and what aspects contributed to those interactions, can help aid designers to improve the transparency embedded in M2 and Collective visualization.

7 R3: SYSTEM DESIGN ELEMENT USABILITY

Understanding which model and visualization combination promoted better usability, R3, is necessary to determine which system design elements promote effective transparency. The associated objective dependent variables were (1) visualization clutter, (2) Euclidean distance, (3) whether an operator was in the middle of an action and completed that action when asked an SA probe question, (4) issued commands, (5) collective and target right-clicks, (6) metrics associated with abandoned targets, (7) the time between the committed state and an issued decide command, and (8) metrics associated with information windows. The specific direct and indirect transparency factors related to R3 are identified in Figure 15. The relationship between the variables and the corresponding hypotheses, as well as the direct and indirect transparency factors, are identified in Table 6. Additional relationships that are not shown in Figure 1, between the variable and the direct or indirect transparency factors, are provided after conducting correlation analyses.

The goal of usability is to design systems that are effective, efficient, safe to use, easy to learn, and are memorable [30]. Good usability is necessary to ensure operators can perceive and
understand the information presented on a visualization and to promote effective interactions. It was hypothesized \( (H_6) \) that the \( M_2 \) model and Collective visualization will promote better usability by being more predictable and explainable. Providing information that is explainable may aid operator comprehension, while predictable information may expedite operator actions. An ideal system will not require constant operator interaction to perform well; therefore, it was hypothesized \( (H_7) \) that operators using \( M_2 \) and the Collective visualization will require fewer interactions.

### 7.1 Metrics and Results

System features were available to the operators to aid task completion. The IA visualization had lower global clutter percentages, which was the percentage of area obstructed by all displayed objects. IA interface operators using the \( M_2 \) model had lower global clutter percentages compared to \( M_3 \). Collective interface operators in general had lower global clutter percentages using \( M_2 \). The IA visualization had lower global clutter percentages in general compared to the Collective visualization. The statistical test details were provided in Section 5.1.

The Euclidean distance (pixels) between the SA probe interest and where the operator was interacting indicated where operators focused their attention. Euclidean distance can be used to assess the effectiveness of the object placements on the display. Larger distances are not ideal, because more time \([18]\) and effort is required to locate and interact with the object. The first requirement of calculating the Euclidean distance was to determine which collective or target was of interest in an SA probe question. Target 3 is the target of interest for the following question: “What collectives are investigating Target 3?” The second requirement was to determine where the operator was interacting with the system (i.e., clicking on the interface). Operators from both visualizations when using \( M_2 \) had shorter Euclidean distances compared to \( M_3 \). Shorter Euclidean distances, however, occurred at all timings for \( SA_3 \) and 15 seconds before asking and during response to an SA probe question for \( SA_0 \) and \( SA_2 \). A significant difference between models occurred for Collective interface operators 15 seconds before asking an \( SA_2 \) probe question. Between-visualizations Mann-Whitney-Wilcoxon tests...
Table 6. Interactions That Influence the Human Operator’s Objective and Subjective Variables, Relative to the Hypotheses, as Well as the Direct and Indirect Transparency Factors (See Figure 15)

| Transparency Factors | Direct | Indirect |
|----------------------|--------|----------|
|                      | Directable | Explainability | Information | Observable | Performance | Understanding | Usability | Control | Effectiveness | Justification | Memorability | Predictability | Reliability | Timing | Subj Vars |
|                      |            |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Obj Vars             |            |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Global Clutter       | $H_6$     |              |             |            |             |              |           |         |              |              |             |               |            |        |          |
| Euclidean Dist.      |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| etween SA            |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Probe Interaction    |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| and Clicks           |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Middle of Action     | $H_7$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| during SA Probe      |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Completed            | $H_6$, $H_7$ | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| Interrupted SA       |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Probe Action         |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Investigate          | $H_7$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| Command              |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Abandon              | $H_7$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| Command              |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Decide               | $H_7$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| Command              |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Collective Rights-   | $H_7$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| clicks               |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Target Rights-       | $H_7$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| clicks               |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Highest Valued       | $H_6$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| Target Abandon       |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Abandon Target       | $H_7$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| Info. Window Open    |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Abandon Request      | $H_6$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| Exceeded Abandon     |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Target               | $H_6$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| Time Commit          | $H_6$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| State and Issued     |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Decide Command       |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Frequency of          | $H_6$, $H_7$ | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| Accessed Target      |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Info. Window         |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Time Target          | $H_6$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| Info. Window Open    |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Time Decision        | $H_6$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| Collective Info.     |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Window Open          |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Time Decision        | $H_6$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| Target Info. Window  |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Open                 |           |             |            |            |             |              |           |         |              |              |             |               |            |        |          |

| Subj Vars            |            |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
|                      |            |             |            |            |             |              |           |         |              |              |             |               |            |        |          |
| Post-trial Command   | $H_6$     |              |             |             |             |              |           |         |              |              |             |               |            |        |          |
| Effectiveness        |           | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
| Post-experiment      | $H_6$     | ✓            | ✓           | ✓           | ✓           | ✓            | ✓         | ✓       | ✓              | ✓              | ✓             | ✓               | ✓           | ✓       |          |
found significant effects using $M_2$ 15 seconds before asking an SA probe question for $SA_0$ ($N = 557$, $U = 43,303$, $p = 0.02$) and $SA_1$ ($N = 273$, $U = 10,577$, $p = 0.05$). A moderate significant effect between visualizations using $M_2$ while being asked an SA probe question was found for $SA_0$ ($N = 464$, $U = 31,052$, $p < 0.01$) and a significant effect for $SA_1$ ($N = 229$, $U = 7,645$, $p = 0.01$). A significant effect between visualizations using $M_2$ during response to an SA probe question was found for $SA_0$ ($N = 499$, $U = 35,029$, $p = 0.02$). Shorter Euclidean distances occurred when IA interface operators used $M_2$, while Collective interface operators had shorter Euclidean distances using $M_3$. The Spearman correlation analysis found a weak correlation between the Euclidean distance of the SA probe’s interest and the operators’ click and SA probe accuracy for the IA interface operators using $M_2$ 15 seconds before asking an SA probe question for $SA_1$ ($r = -0.18$, $p = 0.04$). Weak correlations were found for the Collective interface operators using $M_3$ for $SA_0$ while being asked ($r = 0.14$, $p = 0.04$) and during response to an SA probe question ($r = 0.16$, $p = 0.01$).

The percentage of time an operator was in the middle of an action during an SA probe question identified how often operators were interrupted by the secondary task. Distracted operators may have needed more time to focus their attention on the SA probe question or may have prioritized their interaction over answering the SA probe question. Understanding how distractions may have negatively influenced operator behavior is needed to design the system to promote effective human-collective interactions. Operators using $M_2$ were interrupted less frequently by the SA probe question compared to those using $M_3$ irrespective of the visualization. The percentage of times operators using either visualization were in the middle of an action during an SA probe question for both models for all decision difficulties had min = 0 and max = 100. IA interface operators at $SA_0$, $SA_1$, and $SA_2$ had median = 0 and at $SA_3$ median = 50, while Collective interface

Fig. 16. Euclidean distance between SA probe interest and clicks median (min/max) and Mann-Whitney-Wilcoxon test by SA level between models.
operators had median = 100 at all decision difficulties. The percentage of times operators from both evaluations were in the middle of an action during an SA probe question was significantly different between models for $SA_0$, $SA_1$, and $SA_2$. Between-visualizations Mann-Whitney-Wilcoxon tests identified highly significant effects using $M_2$ for $SA_0$ ($N = 670$, $U = 74,938$, $p < 0.001$), $SA_1$ ($N = 294$, $U = 14,595$, $p < 0.001$), $SA_2$ ($N = 224$, $U = 8,344$, $p < 0.001$), and $SA_3$ ($N = 152$, $U = 3,780$, $p < 0.001$). Highly significant effects between visualizations using $M_3$ were found for $SA_0$ ($N = 672$, $U = 78,456$, $p < 0.001$), $SA_1$ ($N = 253$, $U = 10,944$, $p < 0.001$), $SA_2$ ($N = 252$, $U = 11,172$, $p < 0.001$), and $SA_3$ ($N = 167$, $U = 4,678$, $p < 0.001$). IA interface operators were interrupted less frequently by SA probe questions. The Spearman correlation analysis revealed weak correlations between the middle of an action during an SA probe and SA probe accuracy for the IA interface operators using $M_2$ for $SA_1$ ($r = -0.22$, $p < 0.01$) as well as $M_3$ for $SA_2$ ($r = 0.19$, $p = 0.05$) and $SA_3$ ($r = -0.33$, $p < 0.01$). A weak correlation was revealed for the Collective interface operators using $M_2$ for $SA_3$ ($r = 0.24$, $p = 0.05$).

The percentage of times a participant completed an interrupted SA probe action identified how often operators returned back to their previous task. A system designed to bring an operator back into the loop via engaging prompts, such as the dynamic individual entity behaviors or opacity of support for targets, can mitigate poor human-collective interactions and performance. A system that is easy to remember is desirable to attain optimal operator behavior [29]. IA interface operators using $M_3$ were able to complete 100% of their interrupted actions compared to those using $M_2$, while Collective interface operators using $M_2$ completed approximately 99% of their interrupted actions. The percentage of completed interrupted SA probe actions for both models at all decision difficulties had median = 100 and max = 100. IA interface operators using $M_2$ had median = 100 at all decision difficulties, while Collective interface operators had median = 0. IA interface operators using $M_2$ had median = 0 at $SA_0$, $SA_2$, and $SA_3$ with a median = 100 at $SA_1$. Collective interface operators had median = 0 at $SA_0$, $SA_1$, and $SA_2$ with a median = 100 at $SA_3$. Significant differences existed between models for the IA interface operators for $SA_0$, while no differences existed for the Collective interface operators. Between-visualizations Mann-Whitney-Wilcoxon tests found a significant effect when using $M_3$ for $SA_1$ ($N = 253$, $U = 55,608$, $p = 0.03$). IA interface operators completed more interrupted actions compared those Collective interface operators. No correlations were found between the completed interrupted SA probe actions and SA probe accuracy.

The investigate command permitted increasing a collective’s support for an operator specified target. Additional support for the same target was achieved by reissuing the investigate command.
repeatedly. Generally, operators using $M_2$ and Collective visualization issued fewer investigate commands. The number of investigate commands issued per decision for each combined model and visualization are shown in Figure 17(a). Significant differences were found between models for the number of investigate commands issued per decision for both visualizations at all decision difficulties. Additional between-visualizations Mann-Whitney-Wilcoxon tests identified a moderate significant effect when using $M_2$ for overall decisions ($N = 672, U = 63,866, p < 0.01$) and a highly significant effect for hard decisions ($N = 298, U = 14,066, p < 0.001$). Highly significant effects between visualizations when using $M_3$ were also found for overall ($U = 17,990, p < 0.001$), easy ($N = 396, U = 6,279.5, p < 0.001$), and hard decisions ($N = 276, U = 2,331.5, p < 0.001$).

The abandon command permitted decreasing a collective’s support for a target. The abandon command only needed to be issued once for the collective to ignore a specified target for the duration of a decision. Operators using $M_2$ in general issued fewer abandon commands compared to $M_3$; however, IA interface operators using $M_3$ issued fewer abandon commands for hard decisions. The number of abandon commands issued per decision for both models at all decision difficulties had median = 0 and min = 0. IA interface operators using $M_2$ had max = 8 for overall and hard decisions, while $M_3$ resulted in a max = 2 for overall and easy decisions. The Collective interface operators using both models for overall and hard decisions had max = 2, as well as for easy decisions using $M_3$. Operators using $M_2$ for easy decisions had max = 1. Significant differences were found between models for the number of abandon commands issued per decision with both visualizations for overall and easy decisions. No significant effects between visualizations were found. IA interface operators issued fewer abandon commands compared to Collective interface operators. Collective interface operators using $M_2$ issued fewer abandon commands for overall and hard decisions only.

A collective’s entities stopped exploring targets and moved to the operator selected target when the decide command was issued. A decide request required at least 30% of the collective support for the operator-specified target. Collectives that reached 50% support for a target transitioned into the executing state and the operator was no longer able to influence the collective behavior. IA interface operators using $M_2$ issued fewer decide commands compared to those using $M_3$ or the Collective visualization. The number of decide commands issued per decision for each combined model and visualization are shown in Figure 17(b). Significant differences were found between models for the number of decide commands issued per decision for both visualizations at all decision difficulties. Between-visualizations Mann-Whitney-Wilcoxon tests found highly significant effects using $M_2$ for overall ($N = 672, U = 63,968, p < 0.01$) and easy decisions ($N = 374, U = 21,014, p < 0.001$). A moderately significant effect between visualizations using $M_3$ was found for overall decisions ($U = 57,952, p < 0.01$) and a significant effect existed for easy decisions ($N = 377, U = 19,997, p = 0.05$).

Collective right-clicks and target right-clicks allowed the operator to access the information windows, which provided the number of individual entities in each state and the percentage of collective support for each target. $M_3$ had fewer collective and target right-clicks compared to $M_2$, while the Collective visualization had fewer target right-clicks compared to the IA visualization. The statistical analyses of both metrics were provided in Section 6.1.

Metrics showing how operators used the abandon command were assessed. IA interface operators using $M_3$ abandoned the highest-value target less frequently and had fewer abandoned target information windows open. The statistical analyses of both metrics were provided in Section 6.1. Instances may have occurred when the operator accidentally issued an undesired abandon command or repeatedly issued the abandon command, although targets were abandoned after a single command. The percent of times abandon commands exceeded abandoned targets was examined. Operators using $M_2$ issued fewer repeated abandon commands compared to $M_3$. The percent of times
abandon commands exceeded abandoned targets for each combined model and visualization are shown in Figure 18. Significant differences were found between models for the percent of times abandon commands exceeded abandoned targets with both visualizations for overall and hard decisions. No significant effects between visualizations were found. IA interface operators had fewer repeated abandon commands compared to Collective interface operators. Collective interface operators using $M_3$ had fewer repeated abandon commands for overall and hard decisions.

The time difference between commit state and issued decide command assessed the operator’s ability to predict the collective’s future transition from the committed state (30% support for a target) to executing (50% support for a target). Operators using $M_3$ issued decide commands faster than $M_2$. The time difference between commit state and issued decide command for each combined model and visualization are shown in Figure 19. Significant differences existed between models for the time difference between the commit state and decide command for both visualizations at all decision difficulties. Collective interface operators had smaller time differences between the committed state and issued decide commands compared to those using the IA visualization; however, no significant effects between visualizations were found. IA interface operators using $M_2$ had smaller time differences between the commit state and decide command for hard decisions.

Further analysis of how operators used the collective and target information windows was conducted. The average number of times target information windows were opened per target per decision identified the average frequency at which the information windows were accessed. Operators using $M_3$ in general accessed target information windows less frequently compared to $M_2$. Target information windows were accessed less frequently for operators from both evaluations.
using $M_2$ for easy decisions. The average frequency of an accessed target information window for each combined model and visualization are shown in Figure 20(a). IA interface operators had significantly different average frequencies of accessed target information windows between models for hard decisions, while the Collective interface operators had no significant differences between models. Additional between-visualizations Mann-Whitney-Wilcoxon tests identified a significant effect when using $M_2$ for overall decisions ($N = 619$, $U = 42,857$, $p = 0.02$) and a moderate significant effect for hard decisions ($N = 282$, $U = 7,908.5$, $p < 0.01$). Operators using the Collective visualization accessed target information windows less frequently compared to the IA visualization.

Operators using the target information windows may have used them frequently for short time periods or left them open for long time periods. The average percentage of time a target information window was open per target relative to the decision time for each combined model and visualization are shown in Figure 20(b). IA interface operators using $M_2$ left target information windows open for shorter time periods. Significant differences were found between models for the average time target information windows were open for both visualizations at all decision difficulties; however, no significant effects between visualizations were found.

Particular information windows may have been accessed more frequently for longer time periods. The average percentage of time the decision target information window was open relative to the decision time for each combined model and visualization are shown in Figure 21(a). Operators using $M_2$ left the decision target information window open for shorter periods of time compared to $M_3$. Significant differences were found between models for the time the decision target information window was open for both visualizations at all decision difficulties. Between-visualizations Mann-Whitney-Wilcoxon tests found a highly significant effect using $M_2$ for overall decisions ($N = 672$, $U = 65,102$, $p < 0.001$), as well as significant effects for easy ($N = 374$, $U = 20,114$, $p = 0.01$), and hard decisions ($N = 298$, $U = 12,832$, $p = 0.02$). A moderately significant effect between visualizations using $M_3$ was found for overall decisions ($U = 48,749$, $p < 0.01$), with significant effects for easy ($N = 396$, $U = 17,095$, $p = 0.03$) and hard decisions ($N = 276$, $U = 8,157$, $p = 0.04$). IA interface operators using $M_2$ left the decision target information window open for shorter time periods compared to $M_3$, while the Collective interface operators had shorter time periods using $M_3$.

The average percentage of time the decision collective information window was open relative to the decision time for each combined model and visualization are shown in Figure 21(b). The time the decision collective information window was open was only assessed for the IA evaluation, because the Collective evaluation did not record which collective window was opened or closed. IA interface operators using $M_3$ left the decision collective information window open for shorter periods for easy decisions, and had no significant differences between models for hard decisions. Additional between-visualizations Mann-Whitney-Wilcoxon tests identified a significant effect when using $M_2$ for overall decisions ($N = 374$, $U = 20,114$, $p = 0.01$) and a moderate significant effect for hard decisions ($N = 298$, $U = 12,832$, $p = 0.02$).
time periods compared to $M_2$. IA interface operators had significantly different times for hard decisions.

The post-trial questionnaire assessed the *perceived effectiveness of each request type* (investigate, abandon, and decide), not effective (1) to very effective (7). The investigate, abandon, and decide rankings were ranked higher for operators using $M_3$ when compared to those using $M_2$. Collective interface operators using $M_2$ ranked abandon effectiveness higher. The post-trial effectiveness for each combined model and visualization are shown in Figure 22. Significant differences between models were found in IA interface operator rankings for the decide command and for Collective operator interface rankings for both the abandon and decide commands. Between-visualizations Mann-Whitney-Wilcoxon tests identified a moderate significant effect for the abandon effectiveness using $M_2$ ($N = 56$, $U = 554.5$, $p < 0.01$). IA interface operators using $M_3$ ranked investigate, abandon, and decide effectiveness higher compared to those using the Collective visualization, while Collective interface operators ranked abandon effectiveness higher using $M_2$.

The post-experiment questionnaire assessed the collective’s *responsiveness* to requests, the participants’ *ability* to choose the highest-valued target, and their *understanding* of the collective behavior. IA interface operators using $M_2$ had the best collective responsiveness, operator ability, and understanding versus $M_3$. Collective interface operators ranked the collective’s responsiveness highest using $M_3$, while operator ability and understanding were highest using $M_2$. Statistical test details were provided in the Metrics and Results Section 5.1.
A summary of $R_3$’s results by the hypotheses, with significant results identified, is provided in Table 7. This summary table is intended to facilitate the discussion.

### 7.2 Discussion

The analysis of which combined model and visualization promoted better usability suggests that the IA visualization promoted transparency more effectively than the Collective visualization, while both models had advantages and disadvantages. Operators using the $M_2$ model had less global clutter, due to target information windows being open for less time, smaller Euclidean distances between the interest of a SA probe question and their interaction, were able to complete interrupted actions after answering a SA probe question, and issued fewer abandon and decide commands. $H_6$, which hypothesized that $M_2$ and the Collective visualization will promote better usability by being more predictable and explainable, was not supported. Operators with both interfaces using $M_2$ abandoned the highest-valued target more frequently, which may have occurred due to misunderstanding or poor SA. IA interface operators using $M_2$ were not as timely at predicting when a collective was committed to a target and had the decision collective information window open for a longer time duration (i.e., lower explainability) compared to using $M_3$. The Collective evaluation did not record which collectives were right-clicked on, which impeded the ability to associate right-clicks with a collective; however, a similar reliance on the decision collective information windows may have occurred considering how the Collective interface operators used the target information windows. Future evaluations will validate Collective interface operator usability behavior.

The Collective visualization enabled operators to complete actions prior to an SA probe question and to issue decide commands shortly after a collective was committed to a target. $H_6$ was not supported for the Collective visualization, since more highest-valued targets were abandoned. The continuous display of collective and target information windows promoted higher SA performance for the Collective interface operators using both models. The reliance of the information provided in the pop-up windows suggests that the information was more explainable and reliable than the information provided on the collective icons. Incorporating the numerical percentage of support from the respective collective on a target icon or identifying the most favored target on a collective hub may reduce the reliance of the information windows and simultaneously improve SA by mitigating potential observability issues if the operator must interact with more collectives.

IA interface operators using $M_3$ and Collective interface operators using $M_2$ were able to complete actions that were interrupted by an SA probe question 99% of the time. The memorability of both models and visualizations enabled operators to return to their previous task after answering the SA probe question, because of the required operator engagement ($M_3$) and established expectations of collective behaviors ($M_2$). The predictability of $M_3$ and the Collective visualization justified issuing decide commands shortly after collectives were in a committed state; however, this finding may be biased for $M_3$, because of the required operator influence to achieve the task. The same bias can be attributed to the command effectiveness rankings, which were higher for $M_3$. The IA interface operators’ ability to identify objects on the visualization may have been impeded by displaying all of the individual entities, collective and target icons, and collective and target information windows when the SA probe question inquired about an object further away from the center of the operator’s current attentional focus. Asking SA probe questions about objects at various distances from the operator’s current focal point is necessary to understand how clutter, or moving individual entities, may affect the operator’s ability to identify the object of interest and impact SA performance.

$H_7$, which hypothesized that operators using $M_2$ and the Collective visualization will require fewer interactions, was not supported. $M_2$ enabled fewer commands compared to $M_3$. The IA
Table 7. A Synopsis of R3’s Hypotheses Associated with Significant Results

| Variable | Sub-Variable | Within Model | Between Visualization | Correlation |
|----------|--------------|--------------|-----------------------|-------------|
|          |              | IA | Coll. | M2 | M3 | IA | Coll. |
| Global Clutter | SA0 | H0 | -All | H0 | -All | - | H0 | -B |
| | SA1 | H0 | -All | H0 | -All | - | H0 | -B |
| | SA2 | H0 | -All | H0 | -D | -B,W | - | H0 | -B |
| | SA3 | H0 | -All | H0 | - | - | H0 | -B |
| Euclidean Distance between SA Probe Interests and Clicks | SA0 | H0 | -All | H0 | -All | - | H0 | -W,D |
| | SA1 | H0 | -All | H0 | -B | - | H0 | -B |
| | SA2 | H0 | -All | H0 | -B | - | H0 | -B |
| Middle of Action during SA Probe | SA0 | H0 | H0 | H0 | H0 | - | H0 | -B |
| | SA1 | H0 | H0 | H0 | H0 | - | H0 | -B |
| | SA2 | H0 | H0 | H0 | H0 | - | H0 | -B |
| | SA3 | H0 | H0 | H0 | H0 | - | H0 | -B |
| Completed Interrupted SA Probe Action | SA0 | H0 | H0 | H0 | H0 | - | H0 | - |
| | SA1 | H0 | H0 | H0 | H0 | - | H0 | - |
| Investigate Commands | Overall | H0 | H0 | H0 | H0 | - | H0 | - |
| | Easy | H0 | H0 | H0 | H0 | - | H0 | - |
| | Hard | H0 | H0 | H0 | H0 | - | H0 | - |
| Abandon Commands | Overall | H0 | H0 | H0 | H0 | - | H0 | - |
| | Easy | H0 | H0 | H0 | H0 | - | H0 | - |
| Decide Commands | Overall | H0 | H0 | H0 | H0 | - | H0 | - |
| | Easy | H0 | H0 | H0 | H0 | - | H0 | - |
| | Hard | H0 | H0 | H0 | H0 | - | H0 | - |
| Collective Right-clicks | Overall | H0 | H0 | H0 | H0 | - | H0 | - |
| | Hard | H0 | H0 | H0 | H0 | - | H0 | - |
| Target Right-clicks per | Easy | H0 | H0 | H0 | H0 | - | H0 | - |
| Highest-valued Target | Overall | H0 | H0 | H0 | H0 | - | H0 | - |
| Abandoned | Easy | H0 | H0 | H0 | H0 | - | H0 | - |
| Abandoned Target | Overall | H0 | H0 | H0 | H0 | - | H0 | - |
| Abandon Requests | Overall | H0 | H0 | H0 | H0 | - | H0 | - |
| Exceeded Abandon Targets | Overall | H0 | H0 | H0 | H0 | - | H0 | - |
| Time between Commit State Issued Decide Command | Overall | H0 | H0 | H0 | H0 | - | H0 | - |
| | Easy | H0 | H0 | H0 | H0 | - | H0 | - |
| | Hard | H0 | H0 | H0 | H0 | - | H0 | - |
| Frequency of Accessed Target Information Windows | Overall | H0 | H0 | H0 | H0 | - | H0 | - |
| | Hard | H0 | H0 | H0 | H0 | - | H0 | - |
| Time Target Information Windows Open | Overall | H0 | H0 | H0 | H0 | - | H0 | - |
| | Easy | H0 | H0 | H0 | H0 | - | H0 | - |
| | Hard | H0 | H0 | H0 | H0 | - | H0 | - |
| Time Decision Collective Information Window Open | Overall | H0 | H0 | H0 | H0 | - | H0 | - |
| | Easy | H0 | H0 | H0 | H0 | - | H0 | - |
| | Hard | H0 | H0 | H0 | H0 | - | H0 | - |
| Post-trial | Abandon | H0 | H0 | H0 | H0 | - | H0 | - |
| | Decide | H0 | H0 | H0 | H0 | - | H0 | - |
| | Responsive | H0 | H0 | H0 | H0 | - | H0 | - |

The SA timings are all timings (All), 15 seconds Before asking (B), While being asked (W), and During response (D) to an SA probe question.
visualization enabled fewer abandon and decide commands. Collective interface operators using $M_2$ had better decision-making performance when more investigate commands were issued. Issuing more investigate commands for high-value targets located further from the hub may suggest that the interaction delay embedded in $M_2$, designed to reduce the impacts of environmental bias and improve the success of choosing the ground truth best targets, may have not accommodated operators’ expectations if lower-valued targets were being favored solely because they were closer to the hub. Collective interface operators who issued more commands may have wanted control and directability over the decision-making, which may have occurred due to lower trust or misunderstanding. Investigations are needed to determine if and how trust may influence operators. Operators used different strategies to fulfill the task; however, the most successful strategy promoted more consensus decision-making (i.e., investigate commands), as opposed to prohibiting exploration of targets (i.e., abandon commands). Understanding how operators used commands is necessary to promote effective interactions and produce desired human-collective performance.

The transparency embedded in $M_2$ and the Collective visualization combination did not support the best overall system usability. The IA visualization promoted less clutter, by alleviating the dependence of the collective and target information windows, and promoted fewer interactions. Modifications to both $M_2$ and the Collective visualization must be made to mitigate the highest-valued target being abandoned more frequently, as well as reduce the reliance on the information windows. The assumption that fewer interactions are optimal may not be accurate for all decision difficulties, such as hard decisions. Understanding strategies and justifications for more interactions is necessary to promote transparency that aids operators during particular situations and results in higher human-collective performance.

### 8 R4: SYSTEM DESIGN ELEMENT INFLUENCE ON TEAM PERFORMANCE

Assessing which combined model and visualization promoted better human-collective system performance, $R_4$, is necessary to determine whether the human-collective system transparency aided task completion. An ideal system performs a task quickly, safely, and successfully. The associated objective dependent variables were (1) decision time, (2) selection success rate, and (3) SA probe accuracy. Additional objective metrics were included to support the correlation analyses. The specific direct and indirect transparency factors related to $R_4$ are identified in Figure 23.
Table 8. Interactions That Influence the Human Operator’s Objective and Subjective Variables, Relative to the Hypotheses, as Well as the Direct and Indirect Transparency Factors (See Figure 23)

| Transparency Factors | Direct | Indirect |
|----------------------|--------|----------|
|                      | Explainability | Information | Observable | Performance | Understanding | Usability | Capability | Control | Effectiveness | Efficiency | Justification | Predictability | SA | Timing |
| Directable           | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| Explainability       | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| Information          | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| Observable           | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| Performance          | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| Understanding        | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| Usability            | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| Capability           | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| Control              | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| Effectiveness        | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| Efficiency           | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| Justification        | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| Predictability       | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| SA                   | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |
| Timing               | ✓      | ✓        | ✓          | ✓          | ✓           | ✓         | ✓          | ✓       | ✓              | ✓         | ✓           | ✓             | ✓  | ✓      |

relationship between the variables and the corresponding hypotheses, as well as the direct and indirect transparency factors, are identified in Table 8. Additional relationships between the variable and the direct or indirect transparency factors, not identified in Figure 1, are provided via correlation analyses.

Performance of the human-collective team can be used to assess the effects of the model and visualization transparency on the team’s ability to fulfill tasks. An ideal system design desires high performance rates. It was hypothesized ($H_8$) that the human-collective performance, effectiveness, efficiency, and timing will be better using the $M_2$ model and Collective visualization.

8.1 Metrics and Results

The length of time it took the human-collective team to reach a decision, decision time (minutes), was examined. Collective interface operators using the $M_2$ model had the fastest decision times. The decision time median, min, max, and the Mann-Whitney-Wilcoxon significant effects between
models are shown in Figure 24. Significant differences in decision time were found between models for both visualizations. Between-visualizations Mann-Whitney-Wilcoxon tests identified significant effects when using $M_3$ for overall ($N = 672, U = 50,921, p = 0.03$), easy ($N = 375, U = 15,452, p = 0.04$), and hard decisions ($N = 297, U = 9,521, p = 0.04$). A significant effect between visualizations using $M_3$ was also found for easy decisions ($N = 396, U = 17,376, p = 0.05$).

The selection success rate was the number of correct decisions (the collective moved to the highest-valued target) relative to the total number of decisions. Collective interface operators using $M_3$ had higher selection success rates, while IA interface operators using $M_2$ had higher selection success rates for hard decisions. The selection success rate for both models at all decision difficulties had median = 100, min = 0, and max = 100. Collective interface operators had significant differences in selection success rate between models for overall decisions, while no significant differences between models were found for IA interface operators. Between-visualizations Mann-Whitney-Wilcoxon tests identified highly significant effects when using $M_2$ for overall ($N = 672, U = 64,008, p < 0.001$) and easy decisions ($N = 375, U = 19,845, p < 0.001$), as well as a moderate significant effect for hard decisions ($N = 297, U = 12,761, p < 0.01$). Highly significant effects between visualizations using $M_3$ for overall ($U = 66,360, p < 0.001$), easy ($N = 396, U = 21,662, p < 0.001$), and hard decisions ($N = 276, U = 12,178, p < 0.01$). The Spearman correlation analysis revealed a moderate correlation between decision time and selection success rate using the IA visualization and $M_2$ for easy decisions ($r = -0.42, p < 0.001$) and a weak correlation for overall decisions ($r = -0.27, p < 0.001$). Weak correlations existed using the Collective visualization and $M_2$ for overall ($r = -0.11, p = 0.05$), easy ($r = -0.18, p = 0.02$), and hard decisions ($r = 0.18, p = 0.03$). A weak correlation was found for hard problems using $M_3$ with the IA ($r = 0.32, p < 0.001$) and Collective visualizations ($r = 0.25, p < 0.01$).

The IA and Collective interface operators’ SA probe accuracy when using $M_2$ was higher for SA4, while the IA interface operators had higher SA2 and the Collective interface operators had higher SA0. Collective interface operators had higher SA probe accuracy, compared to the IA interface operators. The detailed statistical analyses were provided in Section 5.1.

Additional Spearman correlation analyses analyzed if any correlations existed between selection success rate and some objective metrics, including collective and target observations and right-clicks, investigate, abandon, and decide commands, as well as the time a decision collective and target information window was open. A weak correlation existed for collective observations using the Collective visualization with $M_2$ for overall decisions ($r = -0.12, p = 0.03$). Weak correlations were found for target observations when using the Collective visualization with $M_3$ for overall ($r = 0.14, p = 0.01$) and hard decisions ($r = 0.16, p = 0.05$). Weak correlations were found for the number
of target right-clicks using the IA visualization with $M_2$ for overall decisions ($r = -0.13, p = 0.02$), and with $M_5$ for overall ($r = 0.1, p = 0.05$) and hard decisions ($r = 0.18, p = 0.03$), as well as when using the Collective visualization with $M_2$ for hard decisions ($r = 0.17, p = 0.04$). Weak correlations were found for the number of investigate commands when using the Collective visualization with $M_2$ for hard decisions ($r = 0.2, p = 0.01$), as well as when using the IA visualization with $M_3$ for easy ($r = -0.16, p = 0.02$) and hard decisions ($r = 0.24, p < 0.01$). Weak correlations were found for the number of abandon commands when using the IA visualization with $M_2$ for easy decisions ($r = -0.19, p < 0.01$) and with $M_3$ for hard decisions ($r = 0.2, p = 0.02$). A weak correlation existed for the number of decide commands using the Collective visualization with $M_3$ for overall decisions ($r = 0.11, p = 0.05$). Weak correlations were found for the time a decision target information window was open when using the Collective visualization with $M_2$ for overall ($r = 0.11, p = 0.04$) and hard decisions ($r = 0.16, p = 0.04$). No significant effects were found for collective right-clicks and the time a decision collective information window was open.

Spearman correlation analyses were conducted to find correlations between selection success rate and weekly hours on a desktop or laptop, mental rotations assessment, and working memory capacity. Weak correlations were found for the weekly desktop or laptop hours for IA interface operators for easy decision using $M_2$ ($r = 0.16, p = 0.02$), and $M_3$ ($r = -0.15, p = 0.04$), as well as Collective interface operators using $M_3$ for hard decisions ($r = 0.17, p = 0.05$). A weak correlation was found for the mental rotations assessment using the IA visualization with $M_3$ for hard decisions ($r = 0.18, p = 0.04$). Weak correlations were found for working memory capacity and easy decisions using the IA visualization with $M_2$ ($r = -0.17, p = 0.02$), and $M_3$ ($r = -0.15, p = 0.04$).

The post-trial performance and understanding questionnaire results assessed the participants’ understanding of the collectives’ behavior and their ability to choose the best target for each decision. The Collective interface operators ranked performance and understanding higher when using $M_2$. The statistical analysis details were provided in Section 6.1.

A summary of $R_4$’s results that show the hypotheses with associated significant results is provided in Table 9. This summary table is intended to facilitate the discussion.

8.2 Discussion

The analysis suggests that the Collective visualization promoted better human-collective system performance; however, the models had advantages and disadvantages. The $M_2$ model promoted faster decision times, while the Collective visualization promoted faster decision times, higher selection success rates, and higher subjective performance. SA performance varied across the model and visualization combinations. $H_6$ hypothesized that the human-collective decision-making performance, effectiveness, efficiency, and timing will be better using $M_2$ with the Collective visualization, which was partially supported. Collective interface operators using $M_2$ had faster decision times; however, $M_3$ enabled higher selection success rates. Embedding transparency into $M_2$ requires (1) balancing control so the operators can positively contribute and direct decision-making, (2) promoting positive human-collective interactions so the operator’s and the collectives’ strengths are maximized, and (3) alleviating the operator’s workload.

Understanding usability and what interactions were used by operators to justify actions that contributed to performance are necessary to identify the most effective and efficient strategies. Operators using $M_2$ issued fewer commands, which was desired to maximize the collectives’ consensus decision-making process; however, more particular interactions, such as investigate commands, resulted in higher selection success rate performance. Requiring operators to influence the task ensured better performance, because they were in-the-loop, versus operators who were supervising the collective behaviors and potentially correcting collectives’ actions towards task
Table 9. A Synopsis of R4’s Hypotheses Associated with Significant Results

| Variable                        | Sub-variable | Within Model | Between Visualization | Correlation |
|---------------------------------|--------------|--------------|-----------------------|-------------|
|                                 |              | IA Coll. M2 M3 | IA Coll. M2 M3 M2 M3 |
| Decision Time                   | Overall      | H8            | H8 H8 H8             | H8 H8      |
|                                 | Easy         | H8            | H8 H8 H8             | H8 H8      |
|                                 | Hard         | H8            | H8 H8 H8             | H8 H8      |
| Decision Time                   | Overall      | H8            | H8 H8 H8             | H8 H8      |
|                                 | Easy         | H8            | H8 H8 H8             | H8 H8      |
|                                 | Hard         | H8            | H8 H8 H8             | H8 H8      |
| Selection Success Rate          | Overall      | H8            | H8 H8 H8             | ————-      |
|                                 | Easy         | H8            | H8 H8 H8             | ————-      |
|                                 | Hard         | H8            | H8 H8 H8             | ————-      |
| SA Probe Accuracy               | SA0          | H8            | H8 H8 H8             | ————-      |
|                                 | SA1          | H8            | H8 H8 H8             | ————-      |
|                                 | SA2          | H8            | H8 H8 H8             | ————-      |
|                                 | SA3          | H8            | H8 H8 H8             | ————-      |
| Collective Observations         | Overall      | H8            | H8 H8 H8             | H8         |
|                                 | Easy         | H8            | H8 H8 H8             | H8         |
|                                 | Hard         | H8            | H8 H8 H8             | H8         |
| Target Observations             | Overall      | H8            | H8 H8 H8             | H8         |
|                                 | Easy         | H8            | H8 H8 H8             | H8         |
|                                 | Hard         | H8            | H8 H8 H8             | H8         |
| Collective Right-clicks         | Overall      | H8            | ————-                | H8 H8      |
|                                 | Hard         | H8            | ————-                | H8 H8      |
| Target Right-clicks per Decision| Overall      | ————-        | H8 H8                | ————-      |
|                                 | Easy         | ————-        | H8 H8                | ————-      |
|                                 | Hard         | ————-        | H8 H8                | ————-      |
| Investigate Commands            | Overall      | H8            | H8 H8 H8             | H8         |
|                                 | Easy         | H8            | H8 H8 H8             | H8         |
|                                 | Hard         | H8            | H8 H8 H8             | H8         |
| Abandon Commands                | Overall      | H8            | H8 H8 H8             | H8         |
|                                 | Easy         | H8            | H8 H8 H8             | H8         |
|                                 | Hard         | H8            | H8 H8 H8             | H8         |
| Decide Commands                 | Overall      | H8            | H8 H8 H8             | H8         |
|                                 | Easy         | H8            | H8 H8 H8             | H8         |
|                                 | Hard         | H8            | H8 H8 H8             | H8         |
| Time Decision Collective        | Overall      | H8            | ————-                | M2 M3 M2 M3|
| Information Window Open         | Hard         | H8            | ————-                | ————-      |
| Time Decision Target Information Window Open | Overall | H8 | H8 | H8 | H8 |
|                                 | Easy         | H8 | H8 | H8 | H8 |
|                                 | Hard         | H8 | H8 | H8 | H8 |
| Mental Rotations Assessment     | SA0          | H8 | H8 | H8 | H8 |
|                                 | SA1          | H8 | H8 | H8 | H8 |
|                                 | SA2          | H8 | H8 | H8 | H8 |
|                                 | Hard         | H8 | H8 | H8 | H8 |
| Working Memory Capacity         | SA0          | H8 | H8 | H8 | H8 |
|                                 | SA1          | H8 | H8 | H8 | H8 |
|                                 | SA2          | H8 | H8 | H8 | H8 |
|                                 | Easy         | H8 | H8 | H8 | H8 |
| Weekly Hours on a Desktop or Laptop | SA0 | H8 | H8 | H8 | H8 |
|                                 | SA1          | H8 | H8 | H8 | H8 |
|                                 | SA2          | H8 | H8 | H8 | H8 |
|                                 | Easy         | H8 | H8 | H8 | H8 |
|                                 | Hard         | H8 | H8 | H8 | H8 |

The SA timings are all timings (All), 15 seconds Before asking (B), While being asked (W), and During response (D) to an SA probe question.
success. Further analysis is required to determine how to improve target selection when using $M_2$. Improvements during training may emphasize the necessity of selecting the highest-value targets. Realistic human-collective scenarios will require high performance with short decision times in uncertain dynamic environments. The design of an effective human-collective system must enable the human-collective team to fulfill primary objectives, without hindering other metrics, such as decision time and accuracy. Devoting more time to ensuring high task performance is a tradeoff [19]. Expedited decisions may have occurred if higher-valued targets were more observable further away from other objects (less clutter), making them more salient, or if impatient operators predicted future collective behaviors and influenced collectives more to make faster decisions. Using target outlines, collective and target information windows, and issuing investigate commands were necessary to fulfill the primary task and can be used to ensure an explainable and usable system. $M_2$ enabled Collective interface operators with different spatial capabilities to perform relatively the same, unlike IA interface operators, specifically those with lower Working Memory Capacity and more weekly desktop or laptop exposure, who had higher selection success rates.

The transparency embedded in the Collective visualization with $M_2$ promoted the fastest decision times; however, modifications are needed to improve other performance metrics. Understanding what interactions contributed to higher performance is necessary to determine what operator strategies are most effective and efficient. $M_2$ subjective performance may have had a consistent negative bias due to learning effects, since it was always presented before $M_3$. Improving the transparency embedded in the Collective visualization to promote better SA performance must be considered. Understanding what IA visualization aspects, such as streamlines between collectives and targets, promoted better SA performance can be emulated in the Collective visualization.

9 DISCUSSION

Assessing how different models and visualizations influenced human-collective robotic system behavior is one contribution of this article. The analysis assessed understanding how the transparency embedded in the combined models and visualizations influenced operators with individual differences (i.e., capabilities), operator comprehension (i.e, capability of understanding), system design element usability (i.e., model and visualization usability), and human-collective performance. This article also determined whether using the best independent model and visualization, derived from two previous results, provided the best transparency. Previous results indicated that the $M_2$ model enabled faster decisions and relied less on operator influence [11], while the Collective visualization provided better transparency [33], because operators with different individual capabilities performed similarly for both tasks, and the human-collective team performed better. $M_2$, independently, did not enable operators with individual differences to perform similarly; however, it did promote fewer interactions and less clutter, which enabled operators to complete interrupted actions, promoted faster decision times, and higher SA performance. The Collective visualization independently enabled operators with different individual differences to perform similarly, promoted higher understanding and SA, enabled operators to complete interrupted actions and issue decide commands shortly after a collective was committed to a target, promoted faster decision times, higher selection success rates, and higher subjective performance. Combined $M_2$ and the Collective visualization promoted lower overall workload, required less physical demand, had fewer investigate commands and target observations (i.e., extra clicks), and enabled the fastest decision time. The different outcomes between the findings in this analysis versus the findings from Cody et al. [11] and Roundtree et al. [33] suggest that transparency cannot be quantified by using the best system design elements, but instead must be quantified by considering how the transparency of the different system design elements interacted, along with the implications of how that system transparency influences human-collective interactions and performance.
Fewer operator interactions was a desired behavior to minimize negative influence on collective behaviors and reduce the reliance on supplementary information; however, operator influence was anticipated to aid the decision-making process and reduce time to complete decisions. This analysis identified positive and negative interactions associated with model and visualization combinations. Collective interface operators relied on visible target information windows more than 25% of the decision time, creating more global clutter, which can hinder effective task performance. Collective interface operators with more clutter were able to answer more SA probe questions correctly and had higher selection success rates. The dependence on visible collective and target information windows may have been influenced by the type of SA probe questions and the visualization not being observable without the supplementary information. Sixteen of 24 SA probe questions depended on collective state and target support numerical values provided in the collective and target information windows. Collective state information was provided via the different color individual entities on the IA visualization and the opacity of the Collective visualization’s hub quadrants, while color and opacity were used to indicate the highest supporting collective on the target icon. The use of opacity may have been ineffective and less salient; however, using different colors to indicate state information may be a possible design modification for the Collective visualization. Experimental design modifications can also be implemented to ensure a more even distribution of SA probe questions that rely on other information, such as the icons, system messages, or collective assignments versus information windows.

The use of target information windows aided Collective interface operators to abandon targets more than 25% of the time. Operators who used the target information windows to justify that a target was abandoned by a collective may have been confused if the reported target support was not equal to zero. Additional abandon commands may have been issued in an attempt to reduce the collective support to zero. IA interface operators may have experienced a similar confusion if they observed entities still travelling to an abandoned target. Implementing design changes, such as showing zero support when an abandon request has been committed or not displaying lost entities after a specific period of time, once a collective hub has moved to a new location, may reduce the number of reissued abandon commands. Collective interface operators using M2 abandoned the highest-valued target more frequently than IA interface operators. Future analysis will determine if the entire target icon must represent the target value more saliently. Opacity levels must also be validated to ensure unique distinctions between low-, medium-, and high-valued targets. Reiterating the task objective, to choose and move each collective to the highest-valued target for each decision, numerous times during training may also help mitigate operator misunderstanding.

Target observations, which were additional target left-clicks that did not influence collective behavior or aid in accessing supplemental information, and interventions, were undesired interactions. IA interface operators may have confused the target integer identifiers with the collective roman numeral identifiers causing target observations. Using distinct identifiers, such as integers versus letters, can potentially reduce the number of observations. IA interface operators’ capability to identify objects far from their current attentional focal point may have been impeded by displaying all of the individual collective entities, collective and target icons, as well as the collective and target information windows. Asking SA probe questions about objects at various distances from the operator’s current focal point is necessary to understand how clutter, or moving individual collective entities, may affect the operator’s ability to identify the SA probe object of interest and answer the question correctly. The use of eye-tracking technology can provide improved insight regarding operator understanding and usability by recording where the operator was looking. Understanding what types of information the operator was potentially perceiving and comprehending, the difficulty of identifying the desired information due to clutter, and the duration of time looking for information will illuminate why operators interacted with the system in a particular way.
Table 10. Additional Human-collective System Design Guidance

| Design Guidance                                                                 |
|---------------------------------------------------------------------------------|
| 1. Provide indicators that identify which particular objects are currently selected, such as the Collective and Target fields in the Collective Request area. |
| 2. Provide control mechanisms that influence the collective consensus decision-making process positively, such as the investigate command. |
| 3. Provide control mechanisms that can undo negative influence, such as cancel assignment. |
| 4. Limit the use of decision-making control mechanism only after a particular certainty value, such as 30% support for a specific target. |
| 5. Limit the amount of times operators can issue particular commands, such as one time for the abandon or decide command. |
| 6. Use underlying intelligent algorithms (e.g., sequential best-of-n decision-making) capable of fulfilling the task without operator influence. |
| 7. Ensure that the underlying intelligent algorithms compensate for environmental biases and other influential factors on the collective processes. |

$M_2$ enabled fewer commands, as expected. Requiring operators to influence the decision-making process ensured better performance, because the operator was in-the-loop, versus operators who were supervising the collective behaviors and potentially correcting actions towards task success. Different strategies were used to fulfill the decision-making task; however, the most successful promoted more consensus decision-making (i.e., investigate commands) versus prohibiting exploitation of particular targets (i.e., abandon commands). The memorability of the models and visualizations enabled operators to come back into-the-loop after answering the SA probe question, because of the required involvement of the operator ($M_3$) and established expectations of collective behaviors ($M_2$). The predictability of $M_3$ and the Collective visualization enabled operators to issue decide commands shortly after collectives were in a committed state. Collective interface operators using $M_3$ reported the best control mechanism responsiveness, which was anticipated due to the amount of operator influence and gained experience using the control mechanisms from $M_2$.

Additional design guidance recommendations, provided in Table 10, were created to expand those from Roundtree et al. [33]. These new recommendations are applicable irrespective of a specific model or visualization type, with a focus on control mechanism and model features. Providing control mechanisms that can influence the collective decision-making process positively are ideal for ensuring task completion. Further investigations are required to determine how to improve the efficacy of control mechanisms, such as abandon, that can negatively influence task completion. Providing control mechanisms to undo undesired abandon commands is recommended. Additional analyses and investigations are needed to verify the effectiveness of the design strategies for real-world scenarios where bandwidth limitations occur. Understanding how information latency and inaccurate collective state information influence human-collective behavior negatively is essential to designing a resilient transparent collective system.

Transparency for human-collective systems can be achieved via different design strategies for specific system design elements and must be assessed holistically by understanding how the different factors impact transparency and are influenced by transparency. The four research questions assessed four categories of transparency factors that contribute to an effective system: (1) different operator individual capabilities, (2) operator comprehension, (3) system usability, and (4) human-collective team performance. Ideal collective systems will enable operators with different individual capabilities to perform relatively the same, promote operator comprehension, be
usable, and promote high human-collective performance. As collective systems grow in complexity (e.g., size, heterogeneity), visualizations that show the individual collective entities will cause perceptual and comprehension challenges, as well as influence operator actions negatively. The same advantageous observation (i.e., dynamically seeing collective behaviors and support) from this analysis may not occur with large collective (>10,000). A collective system designed using the provided guidelines can help promote better transparency and enable effective human-collective teams.

10 CONCLUSION
Designers of human-collective robotic systems continue to debate what models, control mechanisms, and visualizations are needed to provide transparency of collective behaviors to operators. This article evaluates two models, one consensus decision-making model and another that required operator influence to achieve the task, and two visualizations, a traditional and abstract collective representation, for a sequential best-of-n decision-making task with four collectives, each consisting of 200 individual collective entities. The model and visualization transparency were evaluated with respect to how the system design elements impacted the human operators, operator comprehension, usability, and human-collective performance. Both models and visualizations provided transparency differently. The M2 model and Collective visualization combination did not support any of the research questions collectively, but did partially support specific research questions independently. Quantifying system transparency requires evaluating the transparency embedded in the various system design elements, which has not been done in previous analyses, to determine how they interact with one another and influence human-collective interactions and performance. Designers must build collective systems that are effective regardless of how heterogeneous or large the collective size may become, how simple or complex the collective behaviors are, and how real-world use scenarios, such as bandwidth limitations. Models (e.g., intelligent algorithms) that can aid operators to fulfill the sequential decision-making task that require operator influence and collective visualizations that are observable may be more resilient to real-world scenarios and provide transparency to enable effective human-collective teams.

REFERENCES
[1] Gene M. Alarcon, Rose Gamble, Sarah A. Jessup, Charles Walter, Tyler J. Ryan, David W. Wood, and Chris S. Calhoun. 2017. Application of the heuristic-systematic model to computer code trustworthiness: The influence of reputation and transparency. Cogent Psychol. 4 (2017), 1–22.
[2] Ichiro Aoki. 1982. A simulation study on the schooling mechanism in fish. Bull. Jap. Soc. Sci. Fish. 48, 8 (1982), 1081–1088.
[3] C. Chace Ashcraft, Michael A. Goodrich, and Jacob W. Crandall. 2019. Moderating operator influence in human-swarm systems. In IEEE International Conference on Systems, Man and Cybernetics. 4275–4282.
[4] Hasmik Atoyan, Jean-Rémi Duquet, and Jean-Marc Robert. 2006. Trust in new decision aid systems. In Conference on l’Interaction Homme-Machine. 115–122.
[5] Michele Ballerini, Nicola Cabibbo, Raphael Candelier, Andrea Cavagna, Evaristo Cisbani, Irene Giardina, Vivien Lecomte, Alberto Orlandi, Greg Parisi, Andrea Procaccini, Massimiliano Viale, and Vladimir Zdrakovic. 2008. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. Proc. Nat. Acad. Sci. 105, 4 (2008), 1232–1237.
[6] Eric Bonabeau, Marco Dorigo, and Guy Theraulaz. 1999. Swarm Intelligence. Oxford University Press, New York, NY.
[7] Manuele Brambilla, Eliseo Ferrante, Mauro Birattari, and Marco Dorigo. 2013. Swarm intell swarm robotics: A review from the swarm engineering perspectives. Swarm Intell., Springer 7 (2013), 1–41.
[8] Jessie Y. C. Chen, Katelyn Procci, Michael Boyce, Julia L. Wright, Andre Garcia, and Michael J. Barnes. 2014. Situation Awareness-Based Agent Transparency. Technical Report Technical Report ARL-TR-6905. U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.
[9] John P. Chin, Virginia A. Diehl, and Kent L. Norman. 1988. Development of an instrument measuring user satisfaction of the human-computer interface. In Conference on Human Factors in Computing Systems. 213–218.
[10] Jason R. Cody. 2018. *Discrete Consensus Decisions in Human-collective Teams*. Ph.D. Dissertation. Vanderbilt University, Vanderbilt University, Nashville, TN.

[11] Jason R. Cody, Karina A. Roundtree, and Julie A. Adams. 2021. Human-collective collaboration target selection. *Trans. Hum.-robot Interact.* 10, 2 (2021), 1–29.

[12] Iain D. Couzin, Jens Krause, Nigel R. Franks, and Simon A. Levin. 2005. Effective leadership and decision-making in animal groups on the move. *Nature* 433 (2005), 513–516.

[13] Iain D. Couzin, Jens Krause, Richard James, Graeme D. Ruxton, and Nigel R. Franks. 2002. Collective memory and spatial sorting in animal groups. *J. Theor. Biol.* 218 (2002), 1–11.

[14] Mica R. Endsley. 1995. Toward a theory of situation awareness in dynamic systems. *Hum. FACT.: J. Hum. Fact. Ergon. Soc.* 37, 1 (1995), 32–45.

[15] Randall W. Engle. 2002. Working memory capacity as executive attention. *Curr. Direct. Psychol. Sci.* 11, 1 (2002), 19–23.

[16] Eileen B. Entin, Elliott E. Entin, and Daniel Serfaty. 1996. Optimizing aided target-recognition performance. In *Human Factors and Ergonomics Society Annual Meeting*. 233–237.

[17] Maria Fox, Derek Long, and Daniele Magazzeni. 2017. Explainable planning. *arXiv:1709.10256v1*.

[18] Douglas J. Gillan, Kritina Holden, Susan Adam, Marianne Rudisill, and Laura Magee. 1992. How should Fitts’ law be applied to human-computer interaction? *Interact. Comput.* 4, 3 (1992), 291–313.

[19] Russell Golman, David Hagmann, and John H. Miller. 2015. Polya’s bees: A model of decentralized decision-making. *Sci. Adv.* 1, 8 (2015), 1–7.

[20] Deborah M. Gordon. 1999. *Ants at Work: How an Insect Society Is Organized*. Simon and Schuster, Oxford, UK.

[21] Ellen Haas, MaryAnne Fields, Susan Hill, and Christopher Stachowiak. 2009. *Extreme Scalability: Designing Interfaces and Algorithms for Soldier-robotic Swarm Interaction*. Technical Report ARL-TR-4800. U.S. Army Research Laboratory, Aberdeen Proving Ground, MD.

[22] Musad Haque, Waseem Abbas, Abigail Rafter, and Julie A. Adams. 2017. Efficient topological distances and comparable metric ranges. In *IEEE/RSJ International Conference on Intelligent Robotics and Systems*.

[23] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Adv. Psychol.* 1, 3 (1988), 139–183.

[24] Mark St. John, Harvey S. Smallman, and Daniel I. Manes. 2005. Recovery from interruptions to a dynamic monitoring task: The beguiling utility of instant replay. In *Human Factors and Ergonomics Society Annual Meeting*. 473–477.

[25] Shin-Young Jung and Michael A. Goodrich. 2013. Multi-robot perimeter-shaping through mediator-based swarm control. In *International Conference on Advanced Robotics*. 1–6.

[26] René F. Kizilec. 2016. How much information? Effects of transparency on trust in an algorithmic interface. In *Conference on Human Factors in Computing Systems*. 2390–2395.

[27] Andreas Kolling, Steven Nunnally, and Michael Lewis. 2012. Towards human control of robot swarms. In *ACM/IEEE International Conference on Human-robot Interaction*. 89–96.

[28] Andreas Kolling, Katia Sycara, Steven Nunnally, and Michael Lewis. 2013. Human swarm interaction: An experimental study of two types of interaction with foraging swarms. *J. Hum.-robot Interact.* 2 (2013), 103–128.

[29] Jakob Nielsen. 1993. *Usability Engineering*. Academic Press, Indianapolis, IN.

[30] Jenny Preece and Yvonne Rogers. 2007. *Interaction Design: Beyond Human-computer Interaction*. Wiley, West Sussex, England.

[31] Andreavigianni Reina, Gabriele Valentini, Cristian Fernandez-Oto, Marco Dorigo, and Vito Trianni. 2015. A design pattern for decentralised decision making. *PloS One* 10, 10 (2015), 1–18.

[32] Karina A. Roundtree, Jason R. Cody, Jennifer Leaf, H. Onan Demirel, and Julie A. Adams. 2019. Visualization design for human-collective teams. In *Human Factors and Ergonomics Society Annual Meeting*. 417–421.

[33] Karina A. Roundtree, Jason R. Cody, Jennifer Leaf, H. Onan Demirel, and Julie A. Adams. 2021. Human-collective visualization transparency. *Swarm Intell.* 15 (2021), 237–286.

[34] Karina A. Roundtree, Michael A. Goodrich, and Julie A. Adams. 2019. Transparency: Transitioning from human-machine systems to human-swarm systems. *J. Cogn. Eng. Dec. Mak.* 13, 3 (2019), 171–195.

[35] Erol Sahin and William M. Spears. 2005. *Swarm Robotics*. Springer-Verlag, New York, NY.

[36] Jean Scholtz. 2003. Theory and evaluation of human robot interactions. In *Hawaii International Conference on System Sciences*. 10.

[37] Thomas D. Seeley. 2010. *Honeybee Democracy*. Princeton University Press, Princeton, NJ.

[38] Ariana Strandburg-Peshkin, Colin R. Twomey, Nikolai W. F. Bode, Albert B. Kao, Yael Katz, Christos C. Ioannou, Sara B. Rosenthal, Colin J. Torney, Hai Shan Wu, Simon A. Levin, et al. 2013. Visual sensory networks and effective information transfer in animal groups. *Curr. Biol.* 23, 17 (2013), R709–R711.

[39] David J. T. Sumpter. 2006. The principles of collective animal behavior. *Philosop. Trans. Roy. Societ. B* 361 (2006), 5–22.

[40] Gabriele Valentini, Eliseo Ferrante, and Marco Dorigo. 2017. The best-of-n problem in robot swarms: Formalization, state of the art, and novel perspectives. *Curr. Direct. Psychol. Sci.* 4 (2017), 1–9.
[41] Steven G. Vandenberg and Allan R. Kuse. 1978. Mental rotations, a group test of three-dimensional spatial visualization. *Percept. Motor Skills* 47, 2 (1978), 599–604.

[42] Philip Walker, Michael Lewis, and Katia Sycara. 2016. The effect of display type on operator prediction of future swarm states. In *IEEE International Conference on Systems, Man, and Cybernetics*. 2521–2526.

[43] Philip Walker, Steven Nunnally, Michael Lewis, Nilanjan Chakraborty, and Katia Sycara. 2013. Levels of automation for human influence of robot swarms. In *Human Factors and Ergonomics Society Annual Meeting*. 429–433.

[44] Christopher D. Wickens, John D. Lee, Yili Liu, and Sallie E. Gordon Becker. 2004. *An Introduction to Human Factors Engineering*. Pearson Prentice Hall, Upper Saddle River, NJ.

[45] Edward O. Wilson. 1984. The relation between caste ratios and division of labor in the ant genus *Pheidole Hymenopt.*: *Formic. Behav. Ecol. Sociobiol.* 16 (1984), 89–98.

Received September 2020; revised October 2021; accepted November 2021