18th ICCRTS:

"C2 in Underdeveloped, Degraded and Denied Operational Environments"

Paper ID: 041

Comparison between Human and Agent Runs in the ELICIT N2C2M2 Validation Experiments

Topics: 5 - Experimentation, Metrics, and Analysis,
2 - Approaches and Organizations

Authors:

Marco Manso
SAS-085 Member, Portugal

Mary Ruddy
Azigo, USA

Point of contact:

Marco Manso
Rua da Venezuela , n 29, 14 E, 1500 Lisbon, PORTUGAL
email: marco@marcomanso.com

Abstract

Several ELICIT experiments have been conducted in the past with the objective of better understanding and validating NEC tenets and principles, including the NATO NEC C2 Maturity Model (N2C2M2) developed by the NATO SAS-065 Group. These experiments involved human subjects and, more recently, *parameterizable* software agents capable of mimicking humans. Software agents allowed researchers to significantly enlarge the dataset for analysis and validation and obtain a fully controlled – albeit simpler – environment.

We describe in this paper the main differences between ELICIT N2C2M2 validation experiment runs conducted with human subjects from runs conducted with software agents. Across the N2C2M2 *C2 Approaches*, we observed similar trends in the information and cognitive domains, while on information sharing and social interactions (e.g., *what*, *how* and with *whom* to share) we observed a more homogeneous and regular behavior in agents than in humans. Regarding *C2 Approach* performance, we observed similar trends for effectiveness and efficiency results for all C2 Approaches, except that the *Coordinated C2* approach had the best performance in agent runs while *Collaborative C2* had the best performance in human runs.

We conclude with recommendations and future research suggestions for experiments using agents-based ELICIT.

This work received financial support from the Center for Edge Power at the US Naval Postgraduate School.

1 Introduction

This paper is part of an ongoing series of research on the effects of organization structure, allocation of decision rights and network capabilities on mission effectiveness using the ELICIT (Experimental Laboratory for Investigating Collaboration, Information-sharing and Trust). ELICIT was developed by the Command and Control Research Program (CCRP) as a way to perform formal experiments to isolate and quantify C2 factors that influence team effectiveness. Its first versions allowed two organization archetypes (i.e., traditional C2 hierarchy and edge) to be instantiated in the platform using human subjects only (Ruddy, 2007). Subsequently, the ELICIT platform was enhanced to allow broader scenarios to be studied, including the possibility to instantiate a variety of organization structures and models (Ruddy, 2008).

Because human participants frequently behave in uncontrolled ways and are expensive to recruit, software agents were developed to be used in experiments (Ruddy, 2008). The initial ELICIT agents were validated by configuring them so that their behavior closely mapped to the behavior of specific humans in specific runs for a specific experiment (Ruddy, Wynn and McEver, 2009) (Wynn, Nissen and Ruddy, 2010). Since then, additional work has been conducted to make the agents more sophisticated and better able to mimic humans. There are currently over 50 deterministic and stochastic variables that control the ELICIT agent behavior (Ruddy, 2011). Currently, agents are highly configurable, configured to behave deterministically (thus behave as expected), easy to use and inexpensive, but their behavior is simpler and more predictable than human behavior. Humans, on the other hand, can only be influenced to a certain extent, and sometimes don't follow the experiment instructions as closely as an experimenter might wish.

Therefore, software agents are convenient tools to significantly enlarge the ELICIT experimentation datasets and explore preliminary hypotheses, while final validation still requires human subjects as they provide the *gold standard*. Hence, it is crucial to periodically replicate a series of experiments using human and agent participants and subsequently compare them to provide insight into areas where agents validly model human behavior and also areas where agent behavior is not yet sufficiently nuanced to adequately model human behavior for the purposes of a specific group of experiments.

This paper results from an effort to validate the NATO NEC C2 Maturity Model (N2C2M2) developed by NATO SAS-065 using the ELICIT platform. At a first stage, an experimentation baseline using human subjects was established (Manso and B. Manso, 2010) and, at a second stage, an experimentation baseline using software agents was established (Manso, 2012). We present herein a comparison between both baselines and describe their main differences based on the NCW (Network Centric Warfare) theoretical framework (Manso and Nunes, 2008).

This paper is structured as follows: in section 2, we describe the experiments conducted with ELICIT and the obtained datasets; then, in section 3, we present the comparison of results between the human and the agents baselines; and, in section 4, we present the conclusions.

2 Description of Experiments

In this section of the paper, we describe the background theory used for the experimentation work, then we present the design of the experiments and the relevant datasets obtained for human and agent runs, and we conclude by describing the variables that were measured in the platform, from which the comparison will be based.

2.1 Background Theory: NCW and N2C2M2

The experiments theoretical foundation consists in the Network-Centric Warfare (NCW) theory including the NCW tenets[1], NCW Value Chain (SAS-065, 2010), C2 Domains (Alberts and Hayes 2006), C2 Conceptual Reference Model (CRM) (SAS-050, 2006) (Alberts and Hayes, 2006) and C2 Approach Space (SAS-050, 2006) (Alberts and Hayes, 2006). These theories were used to define the experimentation model, its key variables and their interrelations and the experimentation design. The variables of interest for this work cover different portions of the C2 domains, namely, the *Information Domain* (e.g., capability to share, access, display, store, process and protect information), the *Cognitive Domain* (e.g., individual and collective capability to develop high quality awareness) and the *Social Domain* (e.g., C2 processes and the interactions between and among entities).

For the modeling of the C2 Approaches, we resort to the five classes of **Collective C2 Approaches** defined in the N2C2M2, namely: **Conflicted C2**; **De-conflicted C2**; **Coordinated C2**; **Collaborative C2**; and **Edge C2**. These approaches fit into specific regions of the Collective C2 Approach Space[2], as depicted in Figure 1.

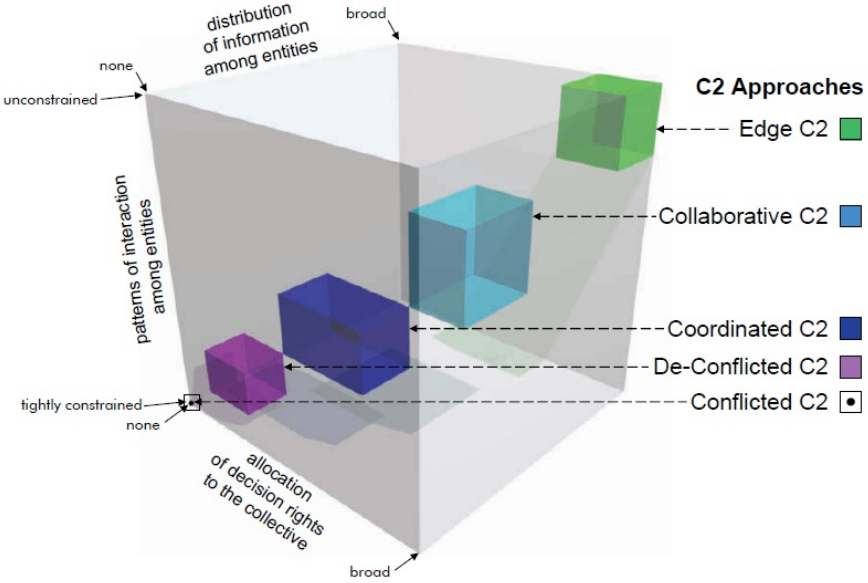


Figure 1 - Collective C2 Approach Space

2.2 Design of Experiments

In order to ensure comparability of results, the same experimentation design was used for human and agent runs. How the C2 approaches are instantiated is explained next, followed by listing the human and agent runs datasets and the list of variables measured.

2.2.1 Instantiation of C2 Approaches

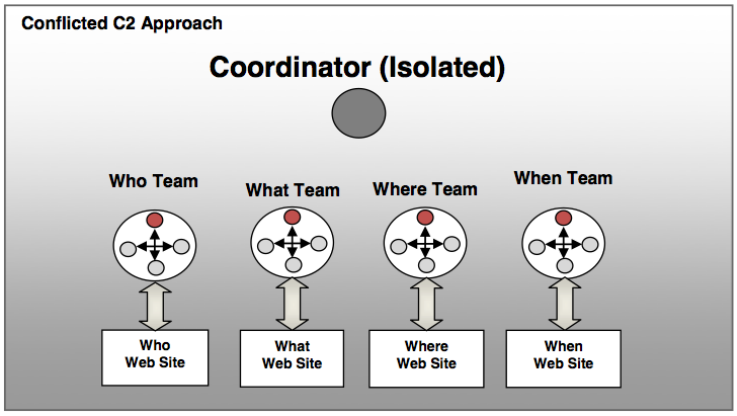
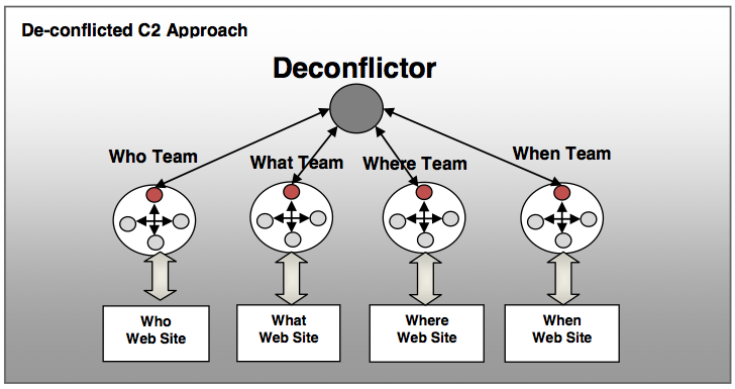
The instantiation of the C2 approaches consists in positioning the collective in specific regions of the Collective C2 Space by manipulating the following dimensions:

[1] Network Centric Warfare Department of Defense Report to Congress. July 2001.

[2] Note that the N2C2M2 deals with the set of entities engaged in a complex endeavor (Alberts and Hayes 2009, 4). Hence, the C2 approach concept is interpreted in the perspective of a 'collective' (i.e., Collective C2). This implies re-interpretation of the dimensions of a Collective C2 approach space (SAS-065 2009, 2) as *allocation of decision rights to the collective* (ADR-C), *patterns of interaction among entities* (PI-C), and *distribution of information among entities* (DI-C).

- **Allocation of decision rights to the collective** (ADR-C): set by (i) determining who has the right to provide identifications that are accounted for mission effectiveness and (ii) setting the agents' parameters so that they do identification actions during a run[3].
- The two interdependent variables "**patterns of interaction among entities**" and "**distribution of information among entities**" are the results of sharing and posting actions and are influenced (in the case of humans) or set (in the case of agents) by the following variables:
  - Individual and Team Characteristics (ITC) (see Table 1).
  - Network characteristics and performance (NCP), set to physically allow or restrict interactions between subjects (i.e., availability of network links between subjects.

In modeling the C2 approaches in ELICIT - including their success criteria and associated characteristics for ADR-C, NCP and ITC, the following acronyms are used: Cross-Team Coordinator or Coordinator (CTC), Deconflictor (Deconf), Coordinator-Facilitator (CF), Team Leader (TL) and Team Member (TM).

C2 Approach Model	Characteristics:
	ADR-C: None. Three roles defined: CTC, TL and TM. Decision rights are allocated to each TL (right to identify in her/his own solution space). NCP: Teams with exclusive access to their website. Non-interoperable (no cross-teams communications). ITC: No sharing of information outside own teams. CTC is isolated. Success Criterion: Each Team pursues independent goals. Success occurs if all TLs find the correct solution to her/his respective problem space. Legend: CTC (grey circle), TL (red circle), TM (light grey circle)
	ADR-C: Established constraints. Three roles defined: Deconf, TL and TM. Decision rights are allocated to each TL (right to identify in her/his own solution space). NCP: Minimum connectivity allowed. Stove-pipe: between TLs and Deconf. Teams have exclusive access to their websites. ITC: Interactions across teams allowed but strictly between each TL and Deconf. Success Criterion: Each Team pursues independent goals for an interdependent problem. Success occurs if all TLs find the correct solution to her/his respective problem space. Legend: Deconf (grey circle), TL (red circle), TM (light grey circle)

[3] Granting decision rights to a subject is not a sufficient condition for her/him to actually take decisions. For example, in a N2C2M2 human run with the Coordinated approach, the Coordinator didn't provide any Identify.

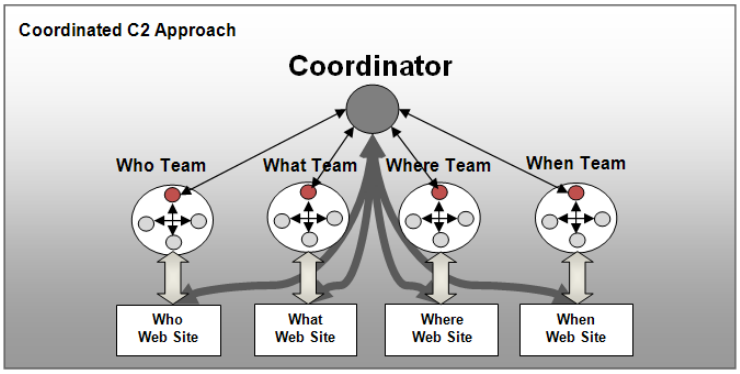
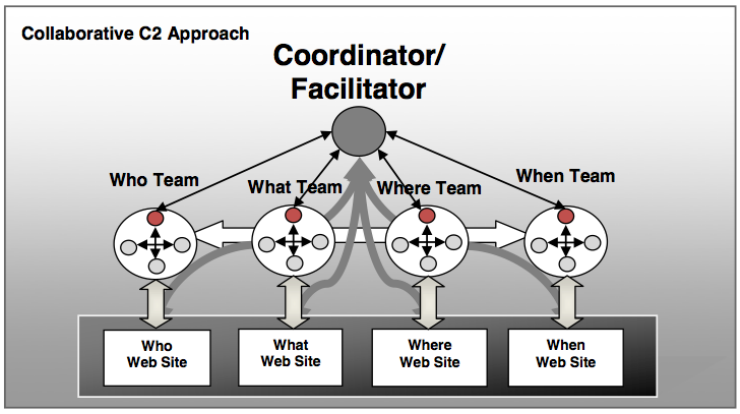
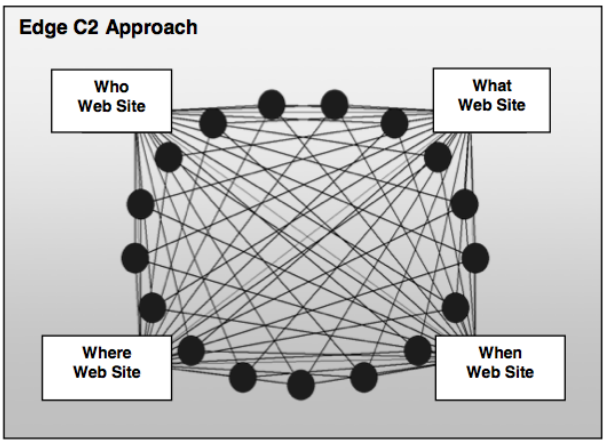
C2 Approach Model	Characteristics:
	ADR-C: Centralized. Three roles defined: CTC, TL and TM. Decision rights are allocated to Coordinator (right to identify in all solution spaces). NCP: Minimum connectivity allowed. Stove-pipe: between TLs and CTC. Teams have exclusive access to their websites and CTC has access to all sites. ITC: Interactions across teams allowed between each TL and CTC. Success Criterion: Organization success depends on the CTC finding the correct solution in all problem spaces. Legend: CTC (grey circle), TL (red circle), TM (light grey circle)
	ADR-C: Distributed and collaborative. Three roles defined: CF, TL and TM. Decision rights allocated to TLs and CF. NCP: Fully connected and interoperable. Existing P2P connectivity between all subjects. Shared team websites. ITC: Interactions allowed between all subjects: CF, TLs and TMs. Success Criterion: Organization success depends on the CF finding the correct solution to all problem spaces OR TLs finding the correct solution to their respective problem space. Legend: CF (grey circle), TL (red circle), TM (light grey circle)
	ADR-C: Fully distributed, not explicit (per individual) and dynamic. One role is pre-defined: TM. TMs choose which part (or parts) of the problem space they work on. NCP: Fully connected and interoperable. Existing P2P connectivity between all individuals. Shared team websites. ITC: interactions allowed between all TMs. Success Criterion: Organization success depends on the individuals' IDs plurality being correct in each problem space. Legend: TM (black circle)

Table 1 – C2 Approach Models and Characteristics

2.2.2 Human Runs Dataset

The human runs dataset results from the N2C2M2 experimentation baseline[4] in (Manso and B. Manso, 2010) and is listed in Table 2. The table includes the run ID, date of the run, filename of the log, C2 Approach that was instantiated and the ELICIT factoid set used. This human runs dataset consisted of 18 runs covering all five C2 Approaches.

[4] Each experimentation baseline includes ELICIT setup files, preparation material (Subject's instructions and ELICIT setup files), subjects' questionnaires and ELICIT logs.

ID	Date	Log File	C2 Approach	Factoid Set
L1-01	13-05-2009	20090513-1053-21812-group_1_CONFLICTED.log	CONFLICTED	4
L1-02	28-05-2009	20090528-1246-32783-group_1_CONFLICTED.log	CONFLICTED	1
L1-03	03-06-2009	20090603-1135-40719-group_1_CONFLICTED.log	CONFLICTED	3
L2-01	29-04-2009	20090429-1308-24386-group_1_B.log	DECONFLICTED	1
L2-02	29-04-2009	20090429-1402-18407-group_1_B.log	DECONFLICTED	4
L2-03	12-05-2009	20090512-1511-12023-group_1_DECONFLICTED.log	DECONFLICTED	1
L2-04	02-06-2009	20090602-1413-19588-group_1_DECONFLICTED.log	DECONFLICTED	3
L3-01	07-05-2009	20090507-1413-26985-group_1_COORDINATED.log	COORDINATED	1
L3-02	13-05-2009	20090513-1142-59475-group_1_COORDINATED.log	COORDINATED	3
L3-03	03-06-2009	20090603-1212-12301-group_1_COORDINATED.log	COORDINATED	2
L3-04	03-06-2009	20090603-1332-06152-group_1_COORDINATED.log	COORDINATED	4
L4-01	06-05-2009	20090506-1133-24903-group_1_COLLABORATIVE.log	COLLABORATIVE	1
L4-02	06-05-2009	20090506-1225-19696-group_1_COLLABORATIVE.log	COLLABORATIVE	4
L4-03	12-05-2009	20090512-1415-19171-group_1_COLLABORATIVE.log	COLLABORATIVE	3
L4-04	02-06-2009	20090602-1515-03801-group_1_COLLABORATIVE.log	COLLABORATIVE	2
L5-01	29-04-2009	20090429-1101-06528-group_1_A.log	EDGE	1
L5-02	29-04-2009	20090429-1152-08894-group_1_A.log	EDGE	4
L5-03	07-05-2009	20090507-1500-37261-group_1_EDGE.log	EDGE	4

Table 2 – N2C2M2 experiment runs

2.2.3 Agent Runs Dataset

The agent runs raw dataset results from the N2C2M2 agent baseline in (Manso, 2011). It consists of 135 runs created by varying different agent archetypes[5] across the organization levels (i.e., top, mid and bottom levels where applicable) in each run. The number of possible combinations is presented in Table 3.

C2 Approach	Agent Type: Top-Level	Agent Type: Mid Level	Agent Type: Bottom-Level	# Possible Combinations*	Run Number
Conflicted C2	1 Coord	4 TLs	12 TMs	27	1 .. 27
De-conflicted C2	1 Deconf	4 TLs	12 TMs	27	28 .. 54
Coordinated C2	1 CTC	4 TLs	12 TMs	27	55 .. 81
Collaborative C2	1 CF	4 TLs	12 TMs	27	82 .. 108
Edge C2	-	-	17 TMs	27**	109 .. 135
			TOTAL	**135**	

* Possible agent types are: (i) baseline, (ii) low-performing and (iii) high-performing.
** Use same combinations of agent types in Edge as for other C2 approaches.

Table 3 – N2C2M2 Agent Baseline

[5] The following agent archetypes were created: average-performing agent, low-performing agent and high-performing agent.

2.3 Measurements

The variables measured from the data extracted from the ELICIT *datalogs* are presented in Table 4. A more detailed explanation and, when applicable, formulas used are presented in 6.1. The measurements are calculated in the same way for human runs and agent runs datasets.

Domain/Category	Variable	Description
Information	Relevant Information Reached (average and per key role)	Relevant factoids reached: - average amount and percentage across all organization - amount per key role (e.g., CTC and TL)
	Shared Relevant Information	Amount of relevant factoids accessible by all subjects. Measured as number and percentage of factoids.
Interactions/Social	Interactions Activity	Average number of interactions (i.e., total shares, posts and pulls) per subject.
	Average Network Reach	Network reach measures the percentage of subjects that a specific subject interacted with. The average network reach is the average value across all organization and is measured here as a percentage.
Cognitive	Time of First Correct ID	The time to first correct and complete identification by any participant.
	Number of Partially Correct IDs	Measures the number of partially correct IDs, that is, identifications that provided a correct solution in any of the problem spaces.
	Cognitive Self-Synchronization (CSSync)	Measures the degree of self-synchronization of the collective in the cognitive domain. See 6.1 for formula and (Manso and Moffat 2010).
MoM	(Mission) Effectiveness	Measures the degree of effectiveness of the organization, based on the criteria defined in Table 1. It is C2 approach dependent.
	(Mission) Time Efficiency	Measures the efficiency of the organization when using time as indication of cost. See 6.1 for formula.
	(Mission) Effort Efficiency	Measures the efficiency of the organization when using effort as indication of cost. See 6.1 for formula.
	Maximum Timeliness	*The time to first correct and complete identification by any participant relative to the time available* (Alberts 2011, 298).

Table 4 – Key Measurements

3 Comparison of Results

In this section, a comparison between the human runs and the agent runs is made. It is organized according to three C2 domains (i.e., Information, Social and Cognitive Domains) established in the NCW Theory and finalizes with the Measures of Merit (MoM).

Note that, in this section and for convenience purposes, each C2 Approach is also identified by a number as follows:

C2 Approach Number	C2 Approach Name
1	Conflicted C2
2	De-Conflicted C2
3	Coordinated C2
4	Collaborative C2
5	Edge C2

3.1 Information Domain

The results of human and agent runs in the information domain are presented in Table 5 and includes (i) **shared information**, (ii) **average information reached** and (iii) **average information reached by the CTC** (subject at the top-level position of the organization). The results obtained across C2 approaches are shown in Figure 2.

C2 Approach Number	Shared Information		Average Information Reached		Average Information Reached (CTC)	
	Human	Agents	Human	Agents	Human	Agents
1	0%	0%	8%	22%	6%	6%
2	0%	0%	48%	25%	35%	29%
3	0%	0%	48%	37%	76%	100%
4	65%	100%	69%	100%	73%	100%
5	75%	100%	75%	100%	-	-

Table 5 - Results of Human and Agent Runs in the Information Domain

Figure 2 - Human and Agent Runs in the Information Domain

- The results for **Shared information** are the same for human and agents for C2 Approaches 1, 2 and 3 (i.e., 0%). For C2 Approaches 4 and 5, while human runs plateaued at 75%, the agent runs always obtained the maximum value (100%).

- For **average information reached** (overall), the overall trend is to increase with more network-centric approaches. Agents were better at sharing more information for C2 Approaches 1, 4 and 5, while humans obtained better scores in C2 Approaches 2 and 3. Looking into the **information reached by the CTC**, the values for humans and agents are similar.

Overall, the trends observed across C2 approaches in the information domain between human and agent runs were similar: more network-centric approaches obtained better scores than less network-enabled ones. The main difference noted between human and agents was that, for C2 Approaches 4 and 5, agent runs achieved 100% shared information while human runs achieved, respectively, 76% and 73% shared information.

3.2 Interactions / Social Domain

The results of human and agent runs in the social domain are presented in Table 6 in terms of (i) **total activity** (number of shares, posts and pulls actions) and (ii) **network reach** (average percentage of agents reached by a given agent). The results obtained across C2 approaches can be visualized in Figure 3.

C2 Approach Number	Interactions Activity		Network Reach	
	Human	Agents	Human	Agents
1	159	42	14%	18%
2	230	43	16%	21%
3	190	46	17%	21%
4	347	118	18%	100%
5	488	118	16%	100%

Table 6 - Results of Human and Agent Runs in the Social Domain

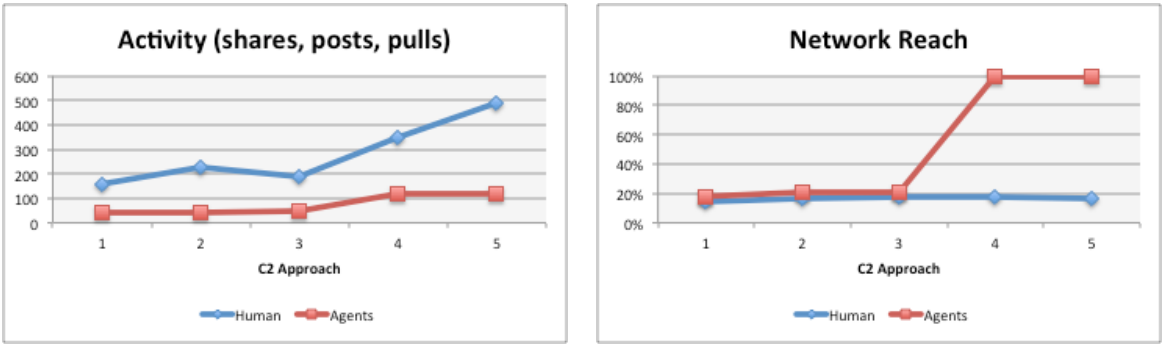
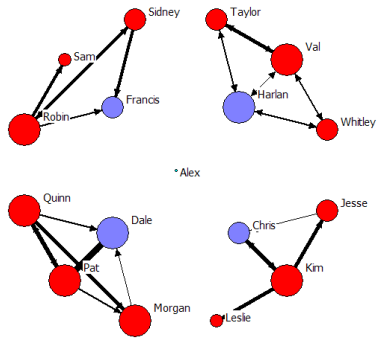
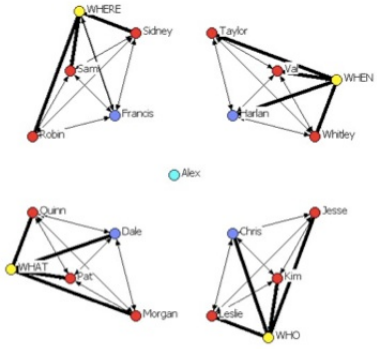
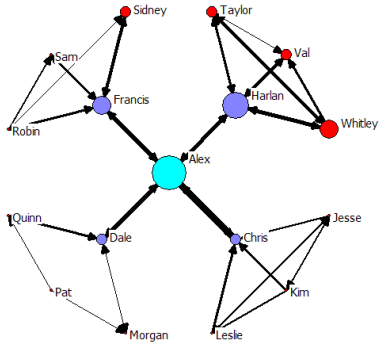
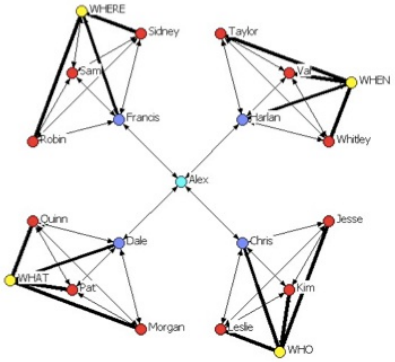
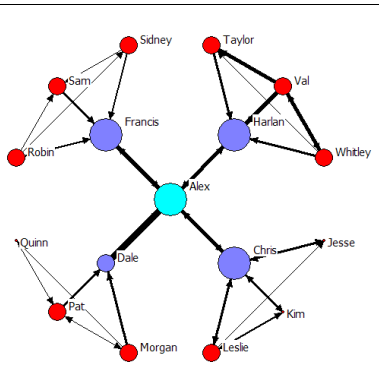
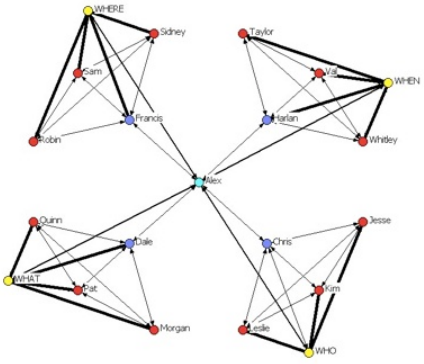


Figure 3 - Human and Agent Runs in the Social Domain

- Agents' **interactions activity** is always below humans'. Both in humans and agents, there is an activity increase when changing from C2 Approach 3 to 4. In humans, the activity continues to increase when changing to Edge C2, while it keeps stable in agents. It should be noted, nonetheless, that agents were more efficient in sharing information per number of actions: for example, with 118 interactions, agents made all information available to the organization while humans spent between 347 and 488 actions (in approaches 4 and 5 respectively) and the **shared information** result was between 65% and 75%.
- The **network reach** values are similar between human and agent runs for C2 Approaches 1, 2 and 3 (about 16%). However, for 4 and 5, agent runs achieve a 100% network reach, but humans do not go above 18%.

The social domain yielded different outcomes for human and agent runs. The agent runs displayed a linear, consistent and undifferentiated behavior (see Table 7). Humans weren't consistent in the way they shared and posted information across the organization. A sociogram per C2 approach for human and agents is presented next (from Manso and B. Manso, 2010, pages 15 to 18). Note that sociograms pertaining to human runs do not include websites. The colors used allow identifying the organization role as follows: CTC, Deconf and CF in cyan; TL in blue; TM in red; websites in yellow (visible in agent runs only).

C2 App.	Human Run[6]	Agent Run	Comments
Conflicted C2			In this C2 approach, all teams are isolated. For the agents, all members interact directly with each other (with same frequency), which is not the case for this human run (humans use websites to share information - not visible in the figure).
De-conflicted C2			Cross-team information sharing is provided via the the Deconf. In this human run, links between Deconf and TLs are the strongest. In agents, there is no differentiation between roles.
Coordinated			The CTC role is created with access to all websites. In this human run, links between the CTC and TLs are the strongest. In agents, there is no differentiation between roles.

[6] The sociograms built for the human runs (Manso and B. Manso 2010) include the additional feature to present the radius of each node proportional to its social activity. That is, nodes with more shares and pulls have a bigger radius than nodes with less shares and pulls.

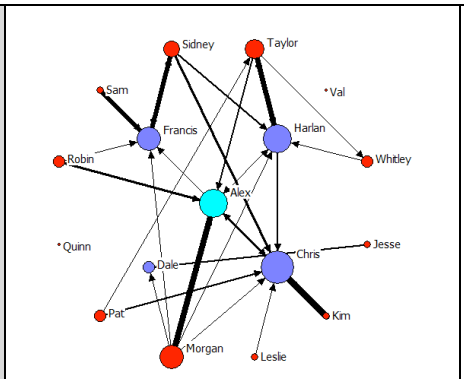
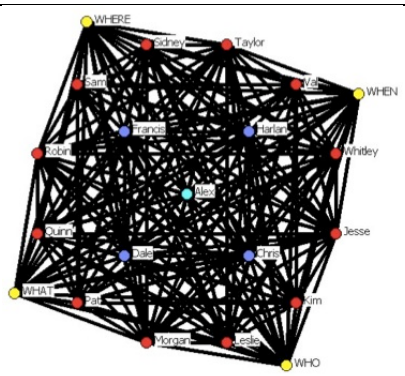
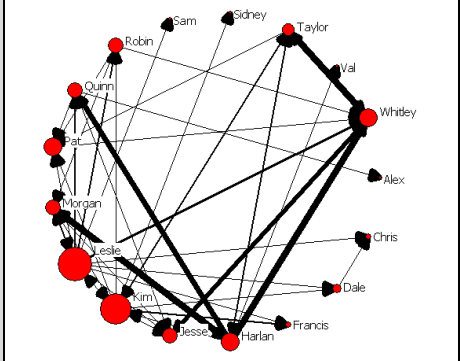
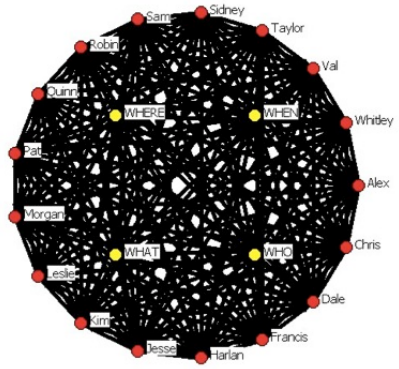
Collaborative C2			A fully connected network is provided in a three-level organisational structure. Humans differentiated their interations based on node role, while agents were fully connected (any member reached all members).
Edge C2			A fully connected network is provided in a flat organisation (no pre-assigned roles). Humans interacted with each other (although websites was their preferred method to post and pull information) while agents were fully connected (any member reached all members).

Table 7 - Human (left) and Agent (right) Sociograms

The social interactions were substantially different between humans and agents, as demonstrated by the measured data and the sociogram examples. While humans display an irregular behaviour (who interacts with whom and how often are always different across runs) - albeit trends do emerge (e.g., emergence of stronger links between key-roles) - agents display a regular behaviour (who interacts with whom and how often is the same across runs in a given C2 approach). It was out of the scope of the agent-based study to replicate the human experiments using a different agent configuration for each subject. Instead, the agents were represented by one of three archetypes. Future experiments should consider using a wider variety of agents. These more varied agents may generate runs whose social interactions are closer to those observed for human runs.

3.3 Cognitive Domain

The results of human and agent runs in the cognitive domain are presented in Table 8 in terms of (i) **number of correct IDs** and (ii) **CSSync**. The results obtained across the C2 approaches can be visualized in Figure 4.

C2 Approach Number	Correct IDs		CSSync	
	Human	Agents	Human	Agents
1	7%	13%	0.05	0.42
2	16%	17%	0.12	0.48
3	11%	15%	0.15	0.52
4	31%	91%	0.34	0.96
5	33%	91%	0.41	0.96

Table 8 - Results of Human and Agents Runs in the Cognitive Domain

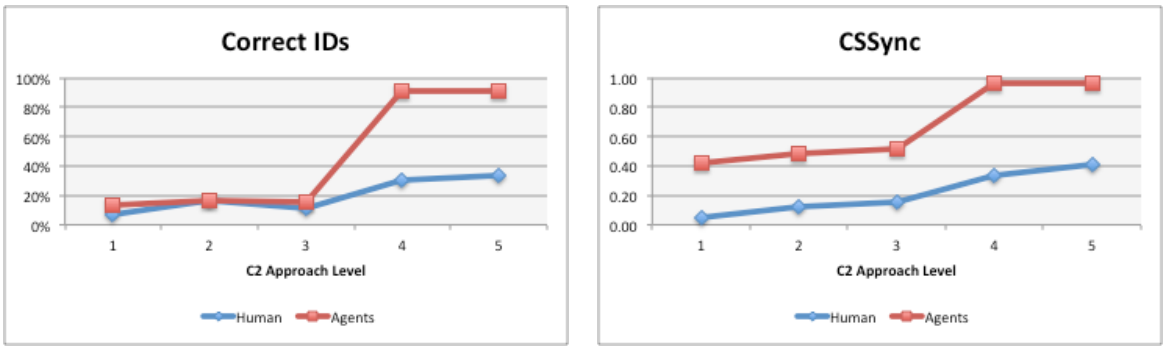


Figure 4 - Human and Agents Runs in the Cognitive Domain

- The number of correct IDs for C2 Approaches 1, 2 and 3 are similar for human and agent runs. For approaches 4 and 5, agent runs reached significantly higher values than for human runs. Achieving success in abELICIT mainly depends in (i) making the necessary information available and (ii) having enough time to process the information. This condition does not always hold for humans.
- The CSSync values are always higher for agent runs than for human runs, Being close to 100% at C2 approaches 4 and 5. CSSync increases with more network-centric approaches.

In what regards the cognitive domain, the observed trends in human and agent runs are similar. However in abELICIT, the solution is determined successfully at all organization levels for Collaborative and Edge as a result of most information being available (and having agents with sufficient information processing performance).

3.4 MoM

The MoM results of human and agent runs are presented in Table 9 in terms of (i) **effectiveness**, (ii) **time-efficiency** and (iii) **effort-efficiency**. The variation trends across the C2 approaches can be visualized in Figure 5.

C2 Approach Number	Effectiveness		Time-Efficiency		Effort-Efficiency	
	Human	Agents	Human*	Agents	Human*	Agents
1	0.31	0.18	0.04	0.04	0.10	0.02
2	0.44	0.30	0.04	0.08	0.13	0.04
3	0.34	0.97	0.06	0.31	0.16	0.29
4	0.80	0.99	0.14	0.32	0.26	0.13
5	0.61	0.95	0.13	0.26	0.12	0.12

* Original values presented in (Manso and B. Manso 2010) were recalculated according to updated efficiency formulas (see Table 14).

Table 9 – MoM Results of Human and Agents Runs

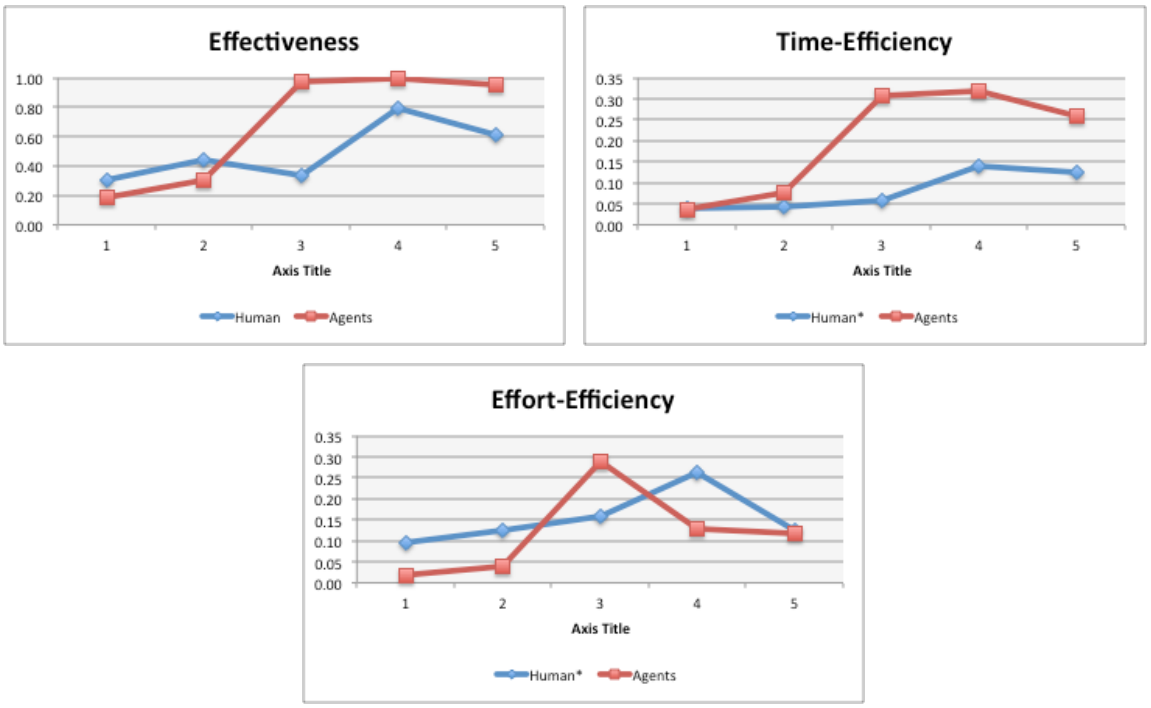


Figure 5 - Trends of MoMs in Human and Agents Runs

In what regards effectiveness:

- Results obtained for C2 Approaches 1 and 2 for human and agent runs are similar (low scores).
- When changing to approach 3, the agent runs have a significant increase while in human runs there is a decrease in effectiveness. Since the agents' had access to the information necessary to determine the solution, the approach was always successful. Humans, on the other hand, were not successful in sharing the necessary information to the CTC or the CTC didn't determine the solution of the problem.
- Moving to approaches 4 and 5, we can see that agents maintain a high effectiveness (above 0.95) while humans increased in 4 but decreased in 5. Once again, the agent's access to all necessary information determined the success of the organizations. The human runs often failed due to lack of access to key information or inability to determine the solution.

Regarding time efficiency, we observe similar trends in human and agent runs (increase for approaches 1 to 4 then decrease), but agents display a significant increase when changing from approach 2 to approach 3.

The effort-efficiency results are significantly different between humans and agents for approaches 3 and 4: The good effectiveness results of agent runs in approach 3 together with the low effort spent (see Table 6 - Coordinated C2 had the second lowest number of interactions) resulted in the best effort-efficiency approach. The human runs were penalized because of the low effectiveness score thus resulting in a low effort-efficiency approach.

Somewhat surprising was the effort-efficiency decrease in approaches 4 and 5 for the agent runs, which is explained due to the high-effort spent in these organizations that, despite the high effectiveness scores, pushed the effort-efficiency to somewhat low values.

For the human runs, approach 4 had highest effort-efficiency but decreased in approach 5, again because of the high-effort spend (post and pull intensive organization).

We note that it is possible to establish optimal strategies for these organizations (e.g., post once and post only)[7] as most of the actions are either redundant or unnecessary (e.g., the same factoid is posted and shared to all team members) but this was not the objective of the experiments. Furthermore, for these experiments, we measured the cost/effort based on the amount of interactions (i.e., share, post, pull and identify), but different organizations may use different criterion as, nowadays, the cost to share/post/pull/process information is becoming irrelevant.

4 Conclusions

Comparing human and agent runs, we observed that similar trends were obtained across information and cognitive domains, but in general agents were more successful than humans in sharing information across all organizations and reach higher values for correct IDs and CSSync in Collaborative and Edge approaches.

However, specifically for the agent runs, our Collaborative model ended located at the top-right of the approach space (i.e., Edge space), but it should be positioned in areas below that region. This could be addressed in the future by modeling the *Collaborative* agents in a more nuanced way. For this work, the decision was made to use this specific Collaborative model in abELICIT to keep backwards comparability with human runs, but a redesign should be considered in future work.

In the social domain, the agents' behavior - which was homogeneous and regular - was significantly different than the human behavior - which was heterogeneous and irregular. All agents followed simple rules to share and post information that resulting in sharing all factoids received to all available team members and websites. Each human, on the other hand, share and post information (i.e., selection of *what*, *how* and *to whom*) based on e.g., personal traits, perceived factoids' relevance, trust and established relations. This human richness and diversity results in unique outcomes in each human trial that has not yet been captured in the ELICIT software agents.

A suggestion for future experiments consists in using abELICIT stochastic parameters related with information sharing to add randomness in the way interactions occur between agents.

Furthermore, looking into the Sensemaking Agent Logic Flow high-level view (Ruddy, Wynn and McEver, 2009) depicted in Figure 6 we can identify some of the high-level components that may be further developed to enrich the agent's social interactions into a more human-like behaviour. The following is recommended for implementation in abELICIT:

- In "Social Processing" add a perceived ranking and trust-level towards other team members that influence how sharing occurs (e.g., share more information to high-valued members than with those that are low-valued members). This ranking is dynamic and built throughout the run.
- In "Information Processing" build a ranking for information that is perceived as having high relevance. The ranking may be a function of internal perceived value and source.
- Add a probability value to share and/or post highly relevant information more than once.
- In "Select Message", give priority to information received from trusted/highly-ranked sources.

[7] We conducted a run in Portugal using the Edge C2 Approach where the human subjects determined as best strategy to post only and post everything. All relevant information was made accessible to the group and they reached an effectiveness score of 50%. While this strategy indeed generates low interactions (and spends a small amount of effort), the lack of collaboration between members (e.g., share key factoids with team) may have limited the increase in shared awareness.

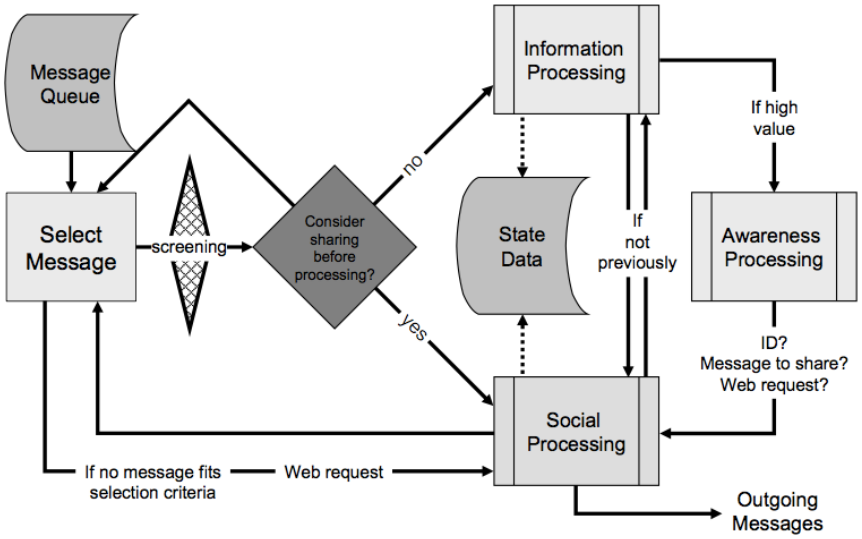


Figure 6 - High-level View of Sensemaking Agent Logic Flow

Regarding the observed MoMs (effectiveness, time-efficiency and effort-efficiency), abELICIT displayed similar trends for agents and humans in all approaches except for Coordinated, which achieved significantly higher values in agent runs than in human runs. For Collaborative and Edge, agent runs also achieved better effectiveness scores than human runs. In abELICIT, agents that have the necessary information accessible are able to determine the solution, a condition that is not valid for most human subjects. Therefore, an improvement is necessary in the agent's "Awareness Processing" that may depend on e.g., arrival of new information, source (trust and ranking) and propensity to change (likelihood to change the ID based on new information received).

5 Bibliography

[1] Alberts, David S., and Richard E. Hayes. *COBP for EXPERIMENTATION*. CCRP Publication Series, 2002.

[2] Alberts, David S., and Richard E. Hayes. *Understanding Command and Control*. CCRP Publication Series, 2006.

[3] Alberts, David S., and Richard E. Hayes. *Planning: Complex Endeavors*. CCRP Publication Series, 2007.

[4] Manso, Marco and Bárbara Manso. *N2C2M2 Experimentation and Validation: Understanding Its C2 Approaches and Implications*. Paper presented at the 15th ICCRTS, Santa Monica, USA, June 22-24, 2010.

[5] Manso, Marco and James Moffat. *Defining and Measuring Cognitive-Entropy and Cognitive Self-Synchronization*. Paper presented at the 16th ICCRTS, Quebec, Canada, June 21–23, 2011.

[6] NATO SAS-050. *Exploring New Command and Control - Concepts and Capabilities - Final Report*. CCRP Publication Series, 2006.

[7] NATO SAS-065. 2010. *NATO NEC C2 Maturity Model*. CCRP Publication Series.

[8] Ruddy, Mary. *ELICIT – The Experimental Laboratory for Investigating Collaboration, Information-sharing and Trust*. Paper presented at the 12th ICCRTS, Newport, USA, 2007.

[9] Ruddy, Mary. *ELICIT 2.2.1 Web-based Software Guide For Human Experiment Directors*. Parity Communications, Inc., Dec 2008. Paper presented at the 14th ICCRTS, Washington DC, USA, 2009.

[10] Ruddy, Mary. Danielle Wynn and Jimmie McEver. *Instantiation of a Sensemaking Agent for use with ELICIT Experimentation*. Paper presented at the 14th ICCRTS, Washington DC, USA, June 15-17 2009

[11] Ruddy, M. *ELICIT 2.5 Software Guide*, CCRP, August, 2011

[12] Ruddy, M. *Pedagogical use of ELICIT for Leadership Training: Survey and Recommendations*. Paper presented at the 16th ICCRTS, Quebec, Canada, June 21–23, 2011.

[13] Wynn, Danielle, Mark Nissen and Mary Ruddy. *The Evolution of C2" Command & Control in Virtual Environments: Tailoring Software Agents to Emulate Specific People*. Paper presented ay the 15th ICCRTS, Santa Monica, USA, June 22-24, 2010.

[14] Network Centric Warfare - Department of Defense Report to Congress. 27 July 2001. http://www.defenselink.mil/cio-nii/docs/pt2_ncw_main.pdf

6 ANNEX

6.1 Data Collection and Measurements

The data collected and measurements obtained from the ELICIT *datalogs* are presented in the next tables.

Name	Value Type	Description
Duration	Number	Duration of a run (in agent's time, measured in Minutes).
Compression factor	Number	Compression of time used to accelerate agent runs (e.g., 0.1 means 1 minute in agents time is 10 minutes in human's time).
Total Shares	Number	Number of shares performed by all members.
Total Posts	Number	Number of posts performed by all members.
Total Pulls	Number	Number of pulls performed by all members.
Total IDs	Number	Number of IDs performed by all members.
List of SenseMaking agent files	Text	Filename of agents file configuration.
Workload	Number	Measured as the number of actions requiring information processing work, that is, number of share received actions and pull actions.

Table 10 – ELICIT measurements: general

Name	Value Type	Description
Relevant facts accessible (number of)	[0..#KES factoids]	Number of K/E/S factoids accessible to organization (currently, the ELICIT platform makes all existing factoids accessible).
Facts accessible (number of)	[0..#factoids]	Number of factoids accessible to organization (currently, the ELICIT platform makes all existing factoids accessible).
Percentage of shared relevant information reached*	[0..100%]	Percentage of K/E/S factoids that were reached by all members.
Relevant information reached per key-role	[0..#KES factoids]	Number of K/E/S factoids that were reached (i.e., shared received or pull) by specific subjects (in key-roles).
Relevant information reached* (mean value)	[0..#KES factoids]	Mean value of K/E/S factoids reached by members of the organization.

#KES factoids = 33, #factoids = 64
(*) reached refers to information that a subject potentially has access to after a (i) pull action or (ii) share received.

Table 11 – ELICIT measurements: information-related

Name	Value Type	Description
Interactions activity (mean value)	Number	Mean value of interaction activities (i.e., number of shared, posts and pulls) per subjects.
Team inward-outward ratio	Number [0..1]	The ratio of inter and intra team interactions (i.e., shares) divided by total number of interactions.
Network reach (mean value)	Percentage	Mean value of the percentage of network reach by subjects.

Table 12 – ELICIT measurements: social-related

Name	Value Type	Description
Number of Partially Correct IDs	[0..4 * nbrSubjects]	Number of partially correct identifications provided by subjects (accounts correct answers in WHO, WHAT, WHERE and WHEN).

Time of First Correct ID	Number	The time to first correct and complete identification by any participant.
CSSync (Cognitive Self-Synchronization)	Number [0..1]	Cognitive self-synchronization value (Marco and Moffat 2011).
CSSync Uncertainty	Number [0..1]	Uncertainty measurement associated with CSSync (Marco and Moffat 2011).

nbrSubjects = 17

Table 13 – ELICIT measurements: awareness-related

Name	Value Type	Description
Effectiveness	[0..1]	Effectiveness score of the organization (Manso and B. Manso 2010)
Time efficiency	Number	Efficiency based on time (Manso and B. Manso 2010) – scaled to 1 hour (3600 seconds): $Efficiency_{time} = Effectiveness_{score}^{2} \times \log_{10}(1+3600/Effectiveness_{time})$
Effort efficiency	Number	Efficiency based on effort (Manso and B. Manso 2011) – scaled to 1000 actions: $Efficiency_{effort} = Effectiveness_{score}^{2} \times \log_{10}(1+1000/totalNumberOfActions)$ Being: totalNumberOfActions = total(shares+pulls+posts+IDs)
MAX timeliness*	Number	*The time to first correct and complete identification by any participant relative to the time available* (Alberts 2011, 298). $MaxTimeliness = 1-\frac{timeFirstCorrectID}{durationOfRun}$

* Concept, variable and formula: (Alberts 2011).

Table 14 – ELICIT measurements: MoMs