Modeling Provenance of Decisions within the Human-Autonomy Team

Crisrael Lucero†, Dr Trung Dong Huynh‡, Dr Douglas S Lange†, Dr Luc Moreau‡

†Space and Naval Warfare Systems Center - Pacific
San Diego, California
UNITED STATES OF AMERICA

‡ King's College London
London, GBR
UNITED KINGDOM

crisrael.lucero@navy.mil, dong.huynh@kcl.ac.uk, doug.lange@navy.mil, luc.moreau@kcl.ac.uk

ABSTRACT

Dynamic mission environments create an always-changing battlespace with many actors. The use of artificial intelligence and autonomy can support the decision-making burden on human operators tasking unmanned vehicles. However, the burden of knowledge and memorization can become overwhelming to bear during complex operations. We introduce information provenance and the capabilities it provides when actors within the human-autonomy team have their actions and decisions modelled by the Provenance Data Model and recorded as provenance data. Issues that are addressed by enabling autonomous command and control systems with information provenance include situation awareness, transparency, and improved decision making aids.

1.0 INTRODUCTION

Provenance is information about entities, activities, agents, and the relationships between these concepts [1]. Provenance explains more than what happened, it also answers how data was manipulated, why, and who was involved in the process. This information should be quite important in military operations and Command and Control (C2), more so when we bring autonomous artificial agents into the decision-making process.

Provenance serves several purposes. Forensically, provenance can help us trace through decisions, understand who made them (human or autonomous agent and under what conditions). Fully recorded provenance can illuminate data dependencies, responsibility flow, and explain why certain actions were taken. During operations, these dependencies are important as well. Dependency tracking is a critical capability of C2, which can lead to explanations and reasoning as to why some decisions cannot be made or need to be updated as both the battlespace and the autonomous actors constantly change status.

As we bring autonomous agents into the C2 process, it is key to find ways of making their decisions transparent to human stakeholders. Some agents will contain decision models that are currently opaque to humans (e.g., artificial neural networks). Tracing the decisions made relative to the information available allows us to find ways of explaining their decisions even if their decision processes are not visible.

This paper describes how we've implemented a provenance recording system that tracks the actions between both humans and several autonomous systems. We also discuss how provenance capabilities can augment human-autonomy teaming and the necessity for provenance to enable explanations and transparency.

2.0 BACKGROUND AND MOTIVATION

The Provenance Data Model (PROV-DM) is a generic data model for provenance that was standardized by the World Wide Web Consortium in 2013. It allows domain and application-specific representations of provenance data to be translated and interchanged between systems [2].

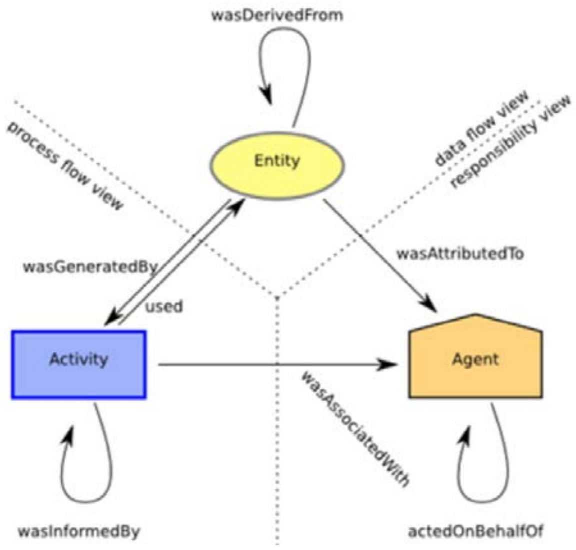

Figure 1: The main PROV concepts. Entities represented as yellow ellipses, activities are blue rectangles, and agents are orange pentagon-houses [2].

The ACM US Public Policy Council published a statement on the importance of algorithmic transparency and accountability, which discusses uncertainty in determining how biased an algorithm may be [3]. There is growing evidence that some algorithms can be opaque, making it impossible to determine whether system outputs are biased or erroneous. The usage of algorithms for automated decision making can potentially be harmful when a system begins to discriminate [4]. AI and autonomous systems have been seen as black boxes, yet these systems should be held at the same standard of ethics as human decision makers. Literature shows a plethora of use cases and methods for using provenance for accountability [5][6] and trust [7][6]. Some ethical aspects of Human-Autonomy Teaming (HAT) can be directly addressed by data provenance. Data provenance brings light on responsibility flow, how data was procured and processed, and explanations for how algorithms evolved data critical to the decision-making process.

Provenance analytics is another growing field of research that provide tools and techniques for provenance graphs.

Anomalies in a workflow execution can be detected by checking for “structural flaws” in its provenance graph, identify graphs whose characteristics are too different from the norms [8]. More recently, it has been shown that, in a number of applications, the topological characteristics of provenance graphs correlate with certain real-world properties of the data or events they describe. For example, predictive models were developed to classify the quality of data created in a crowdsourcing application [9] or to identify instructions from chat message in an alternate-reality game with high levels of accuracy [10].

3.0 PROVENANCE AND HUMAN-AUTONOMY TEAMING

When a system has been enabled with provenance tracking and modeling, different provenance analytics techniques can be used on the collected provenance data to build an autonomous decision support agent. Reference [11] claims that the following attributes represent successful systems: Interactivity, Event and Change Detection, Representation Aiding, Error Detection and Recovery, Information out of Data, and Predictive Capabilities. Additionally, [12] discusses that the following limitations for intelligent decision support systems: emphasis on technology, emphasis on modeling the world with little emphasis on modeling the user, insufficient trust, model incomprehensibility, model scope and model rigidity, and model vulnerability to adversarial attacks. The usage of analyzing and querying provenance for an autonomous agent can both supplement attributes of a successful decision support agent and support the shortcomings that limit their reliability and applicability.

In the context of emergency response, tracking the provenance of decisions was proposed to enable the capability of answering questions on the steps taken in a rescue task, why some target was not rescue, or why there was a delay, etc. [13]. In the same context, the HAC-ER disaster response system demonstrated that tracking provenance of data and decisions across human and agent teams allows for inconsistencies to be automatically detected [14]. Thereby, an autonomous agent monitoring incoming provenance from an active operation was able to alert commanders when new information conflicts with an existing plan sent out to first responders.

Provenance recording in clinical studies facilitates documenting evidences for regulatory submissions and provides auditability for decisions made by the decision support tools used during a study [15].

Our goals for an autonomous system that can query and analyse provenance graphs from systems is to incorporate strong HAT and intelligent decision support agent principles. Within the attributes that represent successful systems, provenance is capable of accomplishing several. The HAC-ER disaster response system showcases the ability to determine event and change detection by querying for and detecting anomalies [14]. Although the visual representations of provenance can be expressed as images, there are different summarization and analysis techniques to aid representation of the scenario. Provenance can also address some of the shortcomings of HAT. Because the PROV-DM is domain-agnostic and can be exchanged between different heterogeneous systems [2], the emphasis can be on both modeling the world and the user. Our specific implementation showcases how provenance is modeling both humans, autonomy, and how they interact with the scenario. Using provenance to build upon trust is an issue that has been thoroughly researched [6][7]. Different analytics techniques have been studied to approach the problem of model incomprehensibility, one such techniques treats provenance graphs as a networks and interprets activities modeled by provenance [9].

The Situation awareness-based agent transparency (SAT) model [16] has been used as a gold standard for modeling how autonomous agents should provide situation awareness and transparency to human operators in order to build trust and accountability. These additional attributes are important to HAT, and being able to convey the different levels of transparency will aid war fighters in developing trust for autonomy, reason over why certain predictions were made, and augment human-in-the loop and human-on-the-loop capabilities. The initial approaches taken to creating a provenance service that not only models provenance, but also queries the provenance generated and provides information about the actions of autonomy to the human operators.

Situation awareness-based Agent Transparency

Level 1: *Agent's current status/actions/plans*
- Purpose: Desire (Goal selection)
- Process: Intentions (Planning/Execution); Progress
- Performance

Level 2: *Agent's reasoning process*
- Reasoning process (Belief/Purpose)
- Environmental & other constraints/affordances

Level 3: *Agent's projections/predictions; uncertainty*
- Projection of future outcomes
- Uncertainty and potential limitations; Likelihood of success/failure
- History of Performance

Figure 2: Situation awareness-based agent transparency (SAT) model [16].

In the following two sections, we discuss how we approach SAT Level 1 and SAT Level 2 by tracking the actions and decisions made by humans and autonomy. We discuss the approach of bringing information of the autonomous agents reasoning processes during violations and how autonomy has changed certain plans using a non-intrusive multimedia narrative.

4.0 ACTIONABLE PROVENANCE WITHIN ALLIED IMPACT

The Intelligent Multi-UxV Planner with Adaptive Collaborative Control Technologies (IMPACT) system is a prototype C2 system that practices different application and platform levels of autonomy [17]. The IMPACT project was an initiative to allow an inversion of the unmanned vehicle staffing ratio; as opposed to several operators controlling a single unmanned vehicle, IMPACT sought to grant a single operator control of several through the use of autonomy [18].

The Allied IMPACT (AIM) system is the joint international effort of several autonomous services working together within the centralized IMPACT system. In a previous study, the IMPACT system was instrumented with the PROV-DM using the Python Prov library [19]. Since then, the core technology has been switched to using the Provenance Templating System (PROV-Templates) to generate provenance [20]. This approach still treats the IMPACT system as a Gray-Box Decision Characterization system and generates expansions of those templates during runtime [21]. PROV-Templates is a declarative approach that enables designers to generate provenance compatible with the PROV standard. Different actions and interactions performed within the AIM system is modeled by individual templates that form the basic placeholder. As bindings are created during application runtime, the templates are instantiated and expanded with real scenario values. The templates are capable of merging together intelligently and provides significant benefits to software engineering [20].

The figure below showcases an example of provenance that has been tracked and modeled based on human interaction with the system and the different autonomous actors that augmented the decision making process. Going through the process flow, a chat event occurred that triggered an autonomous task management agent to discretize the goal and generate a task for the human operator to complete. Based on the different constraints derived by the task manager, a different set of plays is made available by the planning agent to the operator to complete the task. The human operator can choose to engage these plays manually or use a different autonomous agent to carry out the task for them. After the play is selected, the planner will engage the play.

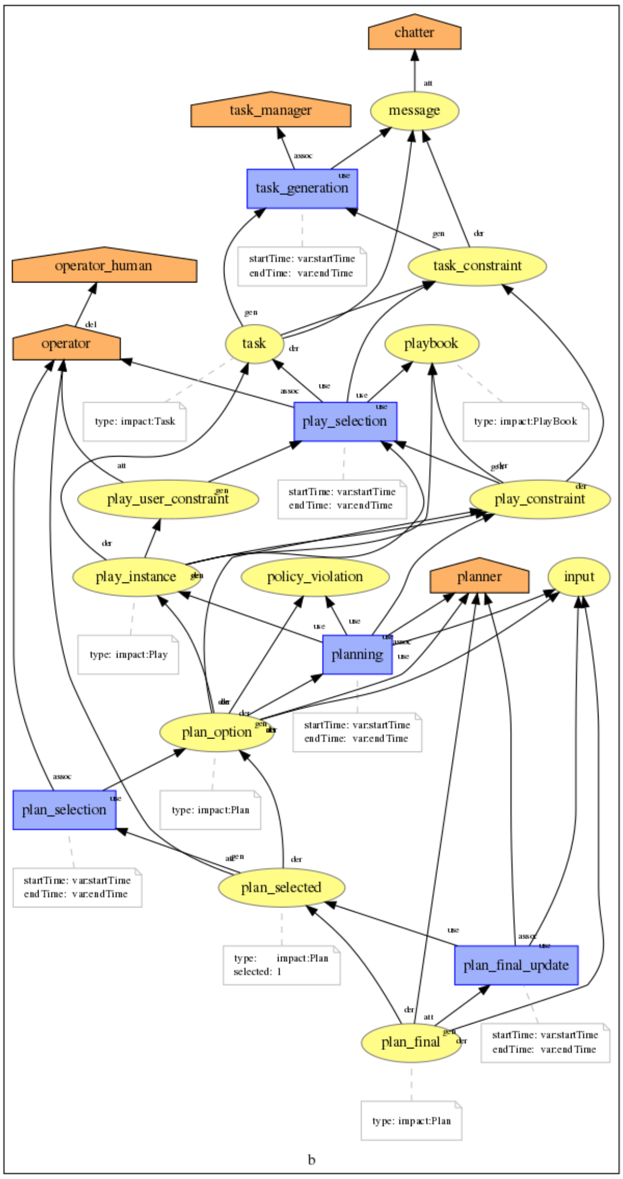

Figure 3: PROV-DM representation of a human play-call. Provenance graphs are to be read left-to-right and top-to-bottom. As provenance graphs grow, visual representations struggle to present information in an organized approach to human users.

Provenance data has also been tracked when autonomous agents interact with each other, whether the human was involved with the process or not. A policy management system autonomously provides warnings to IMPACT when air traffic control policies are violated [19]. In the figure below a policy violation occurs with an ongoing mission and the policy management system notifies IMPACT. The operator agent is not the human, but instead either the Plan Monitor or Task Manager that selects and calls automated plays due to the urgency needed to avoid violating airspace restrictions.

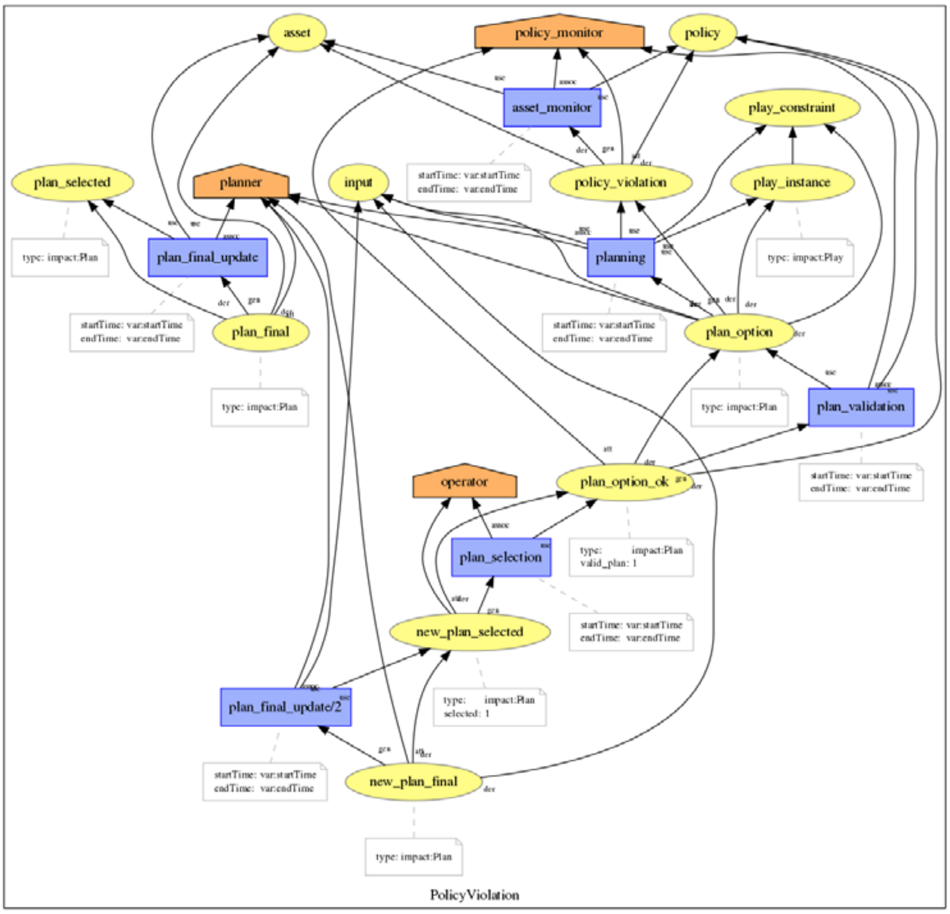

Figure 4: PROV-DM representation of a policy violation and the interactions that occur between the different autonomous systems.

A pilot, proof-of-concept agent has been created for initial testing and basic interfacing with a database of information provenance. The agent allows a human operator to query the provenance and determine which constraints and dependencies are blocking the planner from tasking autonomous vehicles with a play. The provenance agent replies to the human operator in chat a description of which vehicles are capable of accomplishing the play and why those vehicles, based on the application runtime provenance, are unable to accomplish the current task at hand. The next section discusses how we move a step up from using chat-based interactivity and how we've enhanced our agent to do more than dependency and constraint tracking.

5.0 THE NEED FOR EXPLAINABILITY AND TRANSPARENCY

Along with the US ACM statement [3] on the need for algorithmic transparency and accountability, the European Union has passed regulation restricting automated individual decision-making [22]. The regulation effectively creates a "right to explanation" for users so they can ask for explanations of algorithmic decisions that were made about them [23]. With the introduction of explanations generated by autonomous systems using provenance, human operators will not be blindly trusting autonomy. A proposed solution for regulatory mechanisms was to increase the transparency in the data chain [24], which is what data provenance aims to do. The concept of using provenance for explainable artificial intelligence and autonomy has been explored [21], and steps forward can perfect the methodology.

Within the AIM system, the Multimedia Narrative system uses Rhetorical Structure Theory (RST) to provide automated presentations using a virtual human advisor [25]. Work has been done to convert provenance graphs into the RST graph structure to create on-the-fly, automated narratives based on actions autonomy has taken throughout the AIM scenario [26].

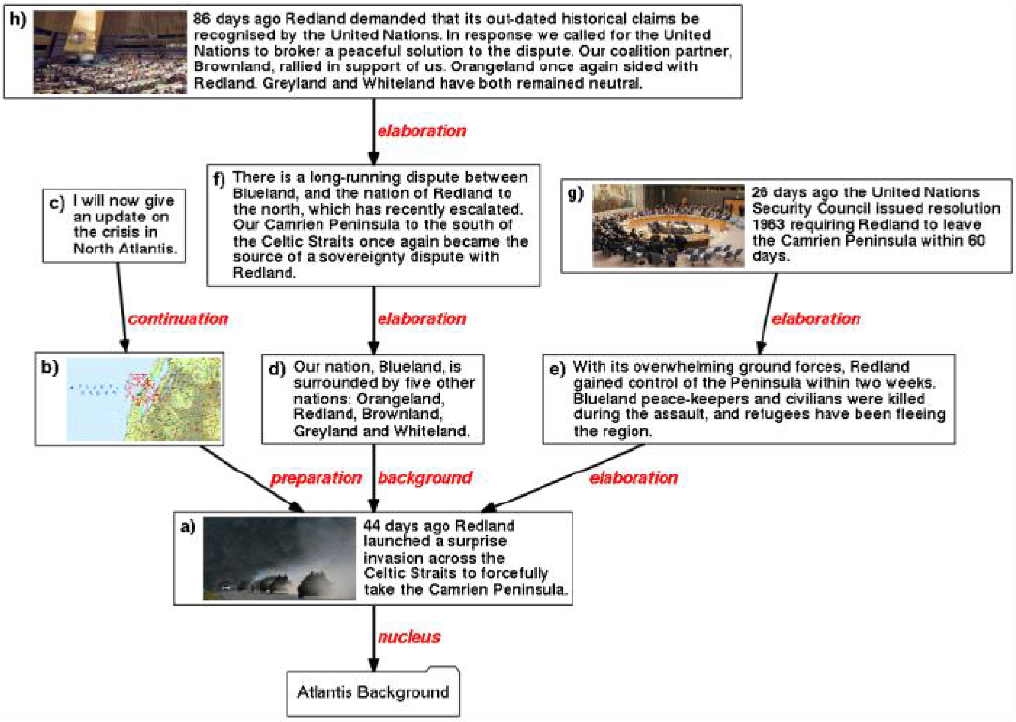

Figure 5: Rhetorical relations assigned to multimedia elements in the 'Military Strikes in Atlantis' scenario [25].

The above figure shows a detailed technical overview of the different rhetoric relations that can be assigned by multimedia elements. The six satellite relations perform a different goal: Preparation establishes narrative context, Elaboration provides more information, Background provides situational context, Conclusion provides a restatement and narrative context, Continuation represents a relation where multiple elements need to be presented together in sequence, and Initialization is a specialized case of preparation that must always be presented with the nucleus [25].

A PROV to RST conversion within the AIM system would allow for the narrative service to select a specific concept (an entity, activity, or agent) from an extracted, segmented piece of the entire application provenance. The specific concept selected acts as the basis of information and rhetoric relations are built based on the provenance relationships (e.g., elaborations and continuations can be built from wasDerivedFrom and wasAttributedTo relationships).

6.0 CONCLUSION AND FUTURE WORK

In this paper we discussed current and ongoing work with enabling and experimenting with provenance within the context of HAT in C2.

The Space and Naval Warfare Systems Center - Pacific along with King's College London and the University of Southern California have begun a research effort to take provenance of C2 information further. Utilizing data analytics and machine learning approaches, the following capabilities are envisioned:

- Learning to characterize and classify the quality of a planning evolution.
- Assigning responsibility for particular decisions to particular agents or teams to illuminate risk to decision makers.
- Summarizing complex composite decisions at appropriate levels of abstraction for informing decision makers.
- Expand upon UML2PROV [27] and attempt to automatically generate PROV-Templates with other modeling languages such as SysML.

With these future capabilities, we hope to further augment our autonomous PROV service with the principles and attributes of HAT systems.

7.0 REFERENCES

[1] Luc Moreau and Paul Groth. Provenance: An Introduction to PROV. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan and Claypool, 2013.

[2] Luc Moreau and Paolo Missier. PROV-DM: The PROV Data Model. *WC3 Recommendation*, 2013. http://www.w3.org/TR/prov-dm/.

[3] Statement on Algorithmic Transparency and Accountability. ACM US Public Policy Council, *USACM Press Releases*, no. January 12, pp. 1—2, 2017.

[4] David Danks and Alex John London. Algorithmic Bias in Autonomous Systems. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 4691—97, 2017.

[5] Daniel J. Weitzner, Harold Abelson, Tim Berners-Lee, Joan Feigenbaum, James Hendler, and Gerald Jay Sussman. Information Accountability. In *Communications of the ACM*, 51, no. 6, pp. 82—87, 2008.

[6] Ryan K.L. Ko, Peter Jagadpramana, Miranda Mowbray, Siani Pearson, Markus Kirchberg, Qianhui Liang, and Bu Sung Lee. TrustCloud: A Framework for Accountability and Trust in Cloud Computing. In *IEEE World Congress on Services*, pp. 584—588, 2011.

[7] Jeremy J. Carroll, Christian Bizer, Pay Hayes, Patrick Strickler. Named Graphs, Provenance and Trust. In *Proceedings of the 14th International Conference on World Wide Web*, pp. 613—622, 2005.

[8] You-Wei Cheah and Beth Plale. Provenance Analysis: Towards Quality Provenance. In *2012 IEEE 8th International Conference on E-Science*, pp. 1—8, 2012.

[9] Trung Dong Huynh, Mark Ebden, Matteo Venanzi, Sarvapali Ramchurn, Stephen Roberts, and Luc Moreau. Interpretation of Crowdsourced Activities Using Provenance Network Analysis. In *The First*

AAAI Conference on Human Computation and Crowdsourcing, pp. 78—85, 2013.

[10] Trung Dong Huynh, Mark Ebden, Joel Fischer, Stephen Roberts, and Luc Moreau. Provenance Networks Analysis. *Data Mining and Knowledge Discovery*, 32(3). Springer US, 2018.

[11] Stephanie Guerlain, Donald E. Brown, and Christina Mastrangelo. Intelligent Decision Support Systems. *IEEE International Conference on Systems, Man, and Cybernetics*, 2000.

[12] Karel van den Bosch and Adelbert Bronkhorst. Human-AI Cooperation to Benefit Military Decision Making. In *Proceedings of the NATO IST-160 Specialist Meeting on Big Data and Artificial Intelligence for Military Decision Making*, 2018.

[13] Iman Naja, Luc Moreau, and Alex Rogers. Provenance of Decisions in Emergency Response Environments. In *Provenance and Annotation of Data and Processes*, pp. 221-30, Springer Berlin, 2010.

[14] Sarvapali D. Ramchurn, Trung Dong Huynh, Feng Wu, Yukki Ikuno, Jack Flann, Luc Moreau, Joel E. Fischer, et al. A Disaster Response System Based on Human-Agent Collectives. In *Journal of Artificial Intelligence Research*, 57, pp. 661-708, 2016.

[15] Vasa Curcin. Embeding Data Provenance into the Learning Health System to Facilitate Reproducible Research. *Learning Health Systems*, 1(2), 2017.

[16] Jessie Y.C. Chen, Shan G. Lakhmani, Kimberly Stowers, Anthony R. Selkowitz, Julia L. Write, and Michael Barnes. Situation Awareness-Based Agent Transparency and Human-Autonomy Teaming Effectiveness. In *Theoretical Issues in Ergonomics Science*, vol. 19, iss. 3, pp. 259—282, 2018.

[17] Robert S. Gutzwiller, Douglas S. Lange, John Reeder, Rob L. Morris, and Olinda Rodas. Human-Computer Collaboration in Adaptive Supervisory Control and Function Allocation. In *7th International Conference on Virtual, Augmented, and Mixed Reality*, 2015.

[18] Mary L. Cummings. Operator Interaction with Centralized versus Decentralized UAV Architectures. In *Handbook of Unmanned Aerial Vehicles*, pp. 977-992, Springer, 2015.

[19] Crisrael Lucero, Braulio Coronado, Eric Gustafson, Douglas S. Lange. Tracking Provenance in Decision Making Between the Human and Autonomy. In *15th International Conference on Engineering Psychology and Cognitive Ergonomics*, 2018.

[20] Luc Moreau, Belfrit V. Batlajery, Trung Dong Huynh, Danius Michaelides, and Heather Packer. A Templating System to Generate Provenance. *IEEE Transactions on Software Engineering*, 44(2), pp. 103—121, 2018.

[21] Crisrael Lucero, Braulio Coronado, Oliver Hui, and Douglas S. Lange. Exploring Explainable Artificial Intelligence and Autonomy through Provenance. *IJCAI-ECAI 2018 2nd Workshop on Explainable Artificial Intelligence*, 2018.

[22] Parliament and the Council of the European Union. General Data Protection Regulation. *Official Journal of the European Union*, 2016.

[23] Bryce Goodman and Seth Flaxman. European Union Regulations on Algorithmic Decision Making and the "Right to Explanation". In *AI Magazine*, vol. 38, no. 3, pp. 50—57, 2017.

[24] Eva Thelisson, Kirtan Padh, and L. Elisa Celis. Regulatory Mechanisms and Algorithms Towards Trust in AI/ML. In *IJCAI 2017 Workshop on Explainable Artificial Intelligence*, 2017.

[25] Steven Wark and Marcin Nowina-Krowicki. Generating Multimedia Narrative for Virtual Humans. In *Proceedings of the Australasian Computer Science Week Multiconference*, 2016.

[26] Marcin Nowina-Krowicki, Peter Argent, Crisrael Lucero, Douglas Lange, and Steven Wark. Now This Changes Everything: Managing Provenance using Multimedia Narrative. *IEEE IRCA Workshop on Robot Teammates Operating in Dynamic, Unstructured Environments*, 2018.

[27] Carlos Sáenz-Adán, Luc Moreau, Beatriz Pérez, Simon Miles, and Francisco J. García-Izquierdo. Automating Provenance Capture in Software Engineering with UML2PROV. In *IPAW'2018: 7th International Provenance and Annotation Workshop*, pp. 1—2, Springer-Verlag Berlin Heidelberg, 2018.