TIPs: Transparency Information Pacts

Phillip Walker, Ugur Kuter, Christopher A. Miller, Helen Wauck, Matthew DeHaven, and Christopher Geib

SIFT, LLC
319 1st Ave N., Suite 400,
Minneapolis, MN 55401, USA
{pwalker,ukuter,cmiller,hwauck,mdehaven,cgeib}@sift.net

Abstract

A key component for building trust in automated systems is improved transparency into the decisions made by the automation and the reasons behind those decisions. Current models of transparency, however, assume that the operator has the time and mental capacity to digest the transparency information the moment it is provided. Often times, this is exactly the moment where the operator is most overloaded, which is why they delegated to automation in the first place. This paper presents the idea of Transparency Information Pacts, or TIPs, as a way to formally represent transparency information and better allow the information to be requested and conveyed back to the human when they are best able to use it. TIPs build off the idea of Lifecycle Transparency presented in (Miller 2021), which encourages the use of other mission phases—pre-mission planning and post-mission debrief—for conveying transparency information where appropriate. Here we present multiple types of TIPs, how they are structured and used, along with illustrative examples.

Introduction

Transparency in automation and AI systems is the ability for the operator to know, or see—as "through glass"—the workings of the machine. Bhaskara, Skinner and Loft's recent review of the literature on transparency (Bhaskara, Skinner, and Loft 2020) says that "In a transparent system, information regarding the agent's actions, decisions, behavior, and intentions is communicated to the operator through an appropriate interface with the aim of improving trust in the system, performance, and operator situation awareness (SA)." Increasing the transparency of a system has generally been shown to improve human SA, trust, and frequently, overall human-machine performance compared to systems which include less or no transparency (Mercado et al. 2016; Lyons and Havig 2014; Ososky et al. 2014; de Visser et al. 2014).

But there is a problem inherent in the concept of transparency as it has frequently been used and researched. Automation is generally installed precisely because human operators do not have time, skill, attentional capacity, or adequate precision to perform the task that the automation does when it does it. And yet, for transparency to function and yield benefit, the human must be able to absorb the transparency information and use it to understand what the automation is doing and determine whether and how to intervene—all in addition to whatever else they need to be doing.

A likely solution to this problem is to take advantage of reduced human task load at other portions of the human-machine integration lifecycle to either transfer transparency information to reduce the need to do so during the highly constrained period of execution or to create conditions so that transfer at execution time will be more efficient and require fewer resources. This might be called taking a LifeCycle Transparency (LCT) perspective (Miller 2021). Echoes of this approach in human-human interactions can be seen in organizational development of standard operating procedures, training in the execution of standard reactions to known or anticipated situations, detailed pre-mission planning and "red teaming", and post-mission debriefing, reviews, and discussions to foster future improvement.

As we have begun researching and designing to support LCT, however, it has become apparent that a missing or under-represented element for lifecycle integration between humans and automation is the existence of the ability to communicate about transparency activities and information. Chen articulated three levels of transparency information having to do with (1) What is going on and what the agent is trying to achieve, (2) Why the agent is doing what it does, and (3) What should the operator expect to happen (Chen et al. 2014). Communication about these constructs, however, requires a shared understanding (and even a shared vocabulary) of (1) states of the world, (2) plans or processes that affect or achieve those states, (3) intentionality to achieve those states or perform those processes, and (4) information presentation capabilities. That is, as human teams do in natural language, we need the ability to talk about these entities outside of the context of execution if we are to gain benefits in execution intervals.

For the most part, items 1-3 are well covered in existing user interfaces. Indeed, the "Playbook" approaches to planning and intentionality expressions described in (Miller and Parasuraman 2007) are, in a sense, designed to be especially efficient means of communicating in well-trained teams. What was missing, however, was an ability to communicate about the conditions under which specific kinds of transparency information might be desired and expected. Plugging that gap is the focus of the work reported here.

This paper will present the concept of Transparency Information Pacts (TIPs), their structure and taxonomy, as well as examples of how TIPs can be integrated into existing Playbook-based systems to improve transparency across the mission lifecycle. TIPs are our answer to what we have identified as as missing piece in the prior use of plays for human-machine teaming. In seeking to define both a human and a machine representation for what plays are and how they can be represented across lifecycle phases in ways useful to both humans and machines, we identified a need to explicitly represent and reason about transparency information, and to do so in ways that would connect pre-mission planning, in-mission execution and information management, and post-mission debriefing phases of operation and their contexts.

Related Work

TIPs are designed to address the need for greater transparency into the reasoning and actions of automated systems during human-in-the-loop operations, while minimizing the cognitive workload required to digest transparency information in-mission. Transparency itself is one manner to establish reliability and predictability in unmanned systems, which enables properly calibrated trust between the human and automation (Bhaskara, Skinner, and Loft 2020; Stowers et al. 2017; Lyons et al. 2019; Smith 2019). Examples on achieving greater system transparency include improved interface design (Kilgore and Voshell 2014), more detailed information readouts (Stowers et al. 2020), improved data fusion (Simpson, Brander, and Portsdown 1995), and explainable artificial intelligence (XAI) (Shin 2021).

Multiple models have been proposed for operationalizing system transparency: Chen et al.'s revised SA Situational Awareness Transparency model (SAT) and Lyons' model of robot-to-human and robot-of-human transparency (Chen et al. 2018; Lyons 2013) are the two primary taxonomies we referenced when designing TIPs. The two models are fairly similar, but the main difference between the two is that Chen et al. include bidirectional transparency (i.e., the human must also convey information to the robot) and Lyons' model adds the concept of "robot-of-human" transparency, which communicates the robot's awareness of factors relating to its human teammates. Robot-to-human, robot-of-human, and bidirectional transparency combined represent mental model synchronization and general awareness of the human and automation of each other.

There is a problem with the current concept of transparency, however, that requires more precise definitions. As mentioned previously, autonomous systems are specifically designed to work when the human operator is not actively monitoring or issuing control inputs to the automated parts of the system. Despite this, laboratory experiments demonstrating benefits for automation often involve operators who are actively monitoring the automation's reasoning and behaviors (Lyons and Guznov 2019; Mercado et al. 2016) and therefore are not representative of most real-world contexts. Other work on trust and transparency with real world scenarios and participants (Lyons et al. 2016; Sadler et al. 2016) either involved limited simulations with novel technologies or, in the case of (Ho et al. 2017), involved pilots reporting on their interactions with novel automation *after* flying. Both the novelty of the automation and the informal experimental settings could provide more monitoring time, attentional capacity, and motivation than would be present in a fully operational setting.

This paper describes a new approach to transparency that will integrate into real-world systems while not assuming there will be available cognitive workload on the part of the operator at any particular point during the mission. Our approach is to create information contracts, or "plays", similar in structure and content to traditional Playbook plays for human-machine teaming (Miller and Parasuraman 2007). Plays represent a shared model of the upcoming task, often in a hierarchical format, for both the human operators and the automation under their control, much in the same way sports teams use plays as an alternative to fully explaining the upcoming plan to every player beforehand every time it is executed (see Figure 1 for an example graphical representation of a play). A play in a human-machine system provides bounds and reasonable defaults for operation, but otherwise leaves the details for the automation to fill in. For instance, one play might task a group of vehicles to ingress from some starting point into a given region. The play might specify maximum route length allowed, whether to proceed with radio silence or not, or no-go zones to avoid. The automation then decides on the exact route. The Playbook approach has previously proven to be an effective method of supervisory control (Fern and Shively 2009).

Similarly, TIPs can provide reasonable bounds and default values for information requests—for instance, if the operator wants an explanation for why a group of unmanned vehicles under their control deviated from their existing course, it could send that request in the form of a TIP that specifies who should respond (all vehicles or just one of them), how deep an explanation is required (just the event that caused the change, or a full chain of reasoning), and a timeline for receiving the response (within 10 minutes, or leave for post-mission debrief). As with plays, these will have sensible defaults so the operator can issue the information request quickly when needed. The configurable timeline for sending and receiving the request allows the operator to make use of displaced transparency (Miller 2021)—in other words, shifting the conveying of transparency to a time when the operator is better able to digest it (e.g., during post-mission debriefing).

TIPs can be tightly integrated with Playbook-based systems, and can themselves be encoded as plays—just ones that require information exchange as the action instead of physical actions in the world. TIPs could also work in concert with existing social AI frameworks (Ehsan et al. 2021; van der Vecht et al. 2018), which aim to move AI interactions and explainability into more human terms. For instance, in (Ehsan et al. 2021) the authors describe how social and organizational contexts are important to consider in the context of XAI systems—the "who", "what", "why", and "when" of a given action by an AI. All of these contexts can be encoded within a TIPs specification.

Finally, previous research has demonstrated that high-performing, well-integrated human teams actually communicate *less* often than other teams (Entin and Serfaty 1999),

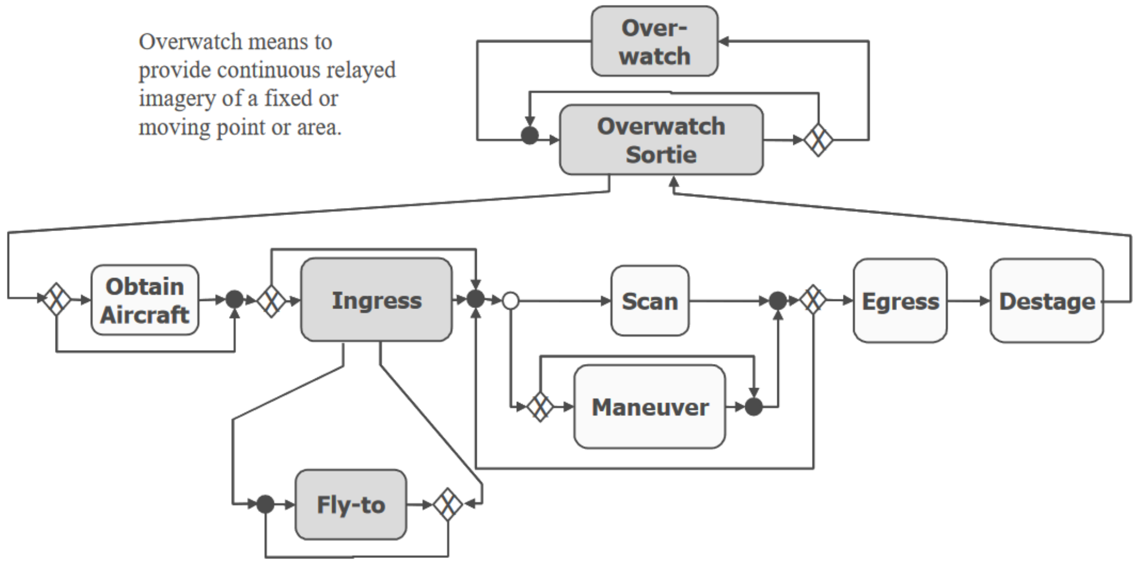

Figure 1: Graphical representation of an Overwatch play from (Miller et al. 2004).

likely because they have better integrated mental models, better communication shortcuts and planning procedures, and the experience of repeated interactions over time, all of which make them better able to predict what other teammates will do. In essence, they have already exhibited displaced transparency in the form of planning and learning from previous interactions, and may already have a TIPs-like communication structure. Our goal with this work is to move toward a formal definition of transparency information exchange in order to improve the function of human-machine teams in the same manner, even when the natural built-up trust from repeated interactions and tight integration is not yet present.

TIPs Taxonomy

The inclusion of TIPs within, or in addition to, a play will dictate when and how information about specific parameters should be conveyed when a play is active. As such, it serves as a contract between the operator, or their management, and the automation about the circumstances in which information will be captured and presented. This enables lifecycle distribution and negotiation about information, as well as the development of expectations about what information will be presented when. This kind of expectation is anticipated to have payoffs for information transference and processing–and associated situation awareness–in terms of both allowing interpretation of the lack of information (e.g., no fuel status warning as long as certain conditions hold) and in terms of setting expectations of the kinds of information to be presented when.

TIP Structure

As plays themselves, TIPs *must* contain the following components, although they may also include other components:

- **Triggers**: Preconditions that must evaluate to true before actions are executed.
- **Actions**: Operations that will execute when all triggers are met.
- **Parameters**: Additional information to take into account when evaluating triggers and executing actions.
- **Caller**: The actor who called, or requested, the TIP/Play.
- **Recipient**: The actor on whom the TIP/Play is called.

Additionally, it may be useful to codify a **topic** for the TIP. The topic refers to the action, state variable, or piece of information with which the TIP is concerned. For instance, a TIP requesting an unmanned vehicle notify the human operator when battery level drops below 30% would have battery level as the topic. Similarly, a TIP requesting the vehicle notify the operator when a previously called action play begins execution would have that play as its topic.

Triggers are defined as any Boolean expression over the state variables available to the TIP callee. TIP actions are distinct from standard play actions in that they satisfy requests for information; they are not physical actions in the real world (although physical actions may need to occur to satisfy the information request).

TIPs can theoretically include multiple callers—the actor who made the explicit TIP call, as well as other actors whom the caller would like to receive the information from the TIP request (e.g, the operator may call a TIP on a single UAV to alert them *and* all other human operators when a specific event occurs). Similarly, there may be multiple callees, for instance when the human wants to receive an alert from *any* UV upon sighting an unknown aircraft.

TIP Types and Classification

We have codified six types of TIPs that we believe cover the range of information requests during a human-automation interaction scenario:

- **Alert**: Inform when a certain trigger has occurred and can have varying degrees of urgency. An ALERT TIP might be called on a UV to request it notify the human operator when the next stage of its mission phase or ongoing play begins (e.g., "Ingress complete, beginning loiter while waiting for further instruction.").
- **Monitor**: Provide continuous updates about a target state variable. A MONITOR TIP might be called on a UV asking it to track a sighted unknown UV, providing continuous estimates of it's position within the area of operations.
- **Request**: Retrieve and return a specific piece of information to the callee. A REQUEST TIP might be called on a UV to request diagnostic information about its onboard camera for analysis by the human to determine the cause of a malfunction.
- **Explain**: Provide justification for an action. This could be associated with a specific action, or all actions within a particular time period or area of operation. An EXPLAIN TIP might be called on a UV asking it to justify why it altered course when the caller was expecting it to maintain heading.
- **Evaluate**: Assess the impact of a given event on the current plan. An EVALUATE TIP might be called when unexpected rain is forecast for part of the area of operations and the human operator wants to know the impact modifying existing no-fly zones to avoid the rain will have on nearby UV plans.
- **Prepare**: Prepare a new plan or play in response to some supplied information. A PREPARE TIP might be called when an operator wants to force a UV to replan an ingress route after the original route closes unexpectedly.

Categorizing TIPs characteristics in different ways can help us understand the types of information different TIPs might require, the different modes of presentation in a human-readable interface, and other essential features required by different TIP structures. Two important characteristics of any TIP to keep in mind when authoring them and designing interfaces are **what** the TIP is concerned with, and **when** is the TIP relevant.

What the TIP caller is requesting transparency information about can be one of two categories: managed (an asset or feature of the world under the operator or their team's control) and unmanaged (not under their control). This distinction is important because it impacts how the information is gathered and how quickly, as well as how reliable one can expect the information to be. An operator calling a TIP on a platform under their control, about a state variable concerning that platform, can be expected to result in a quick answer that is likely accurate, as the platform merely needs to inspect its own state or query its own sensors. An operator calling a TIP requesting information about an unmanaged platform or world state not under their control, however, often requires extra action in the world on the part of the callee, as well as inference about an external state that is not fully observable.

When the TIP is active, and when the event with which the TIP is concerned occurred, also significantly impacts how the calling human specifies the TIP and how it is visualized in their display. The event the TIP is focused on can be in the past, present, or future; and the TIP can be called and executed immediately (present) or later (future). Certain TIPs are only applicable at certain times—for instance, ALERT TIPs can be called immediately or planned for the future, but will always be about future (potential) events. Simiarly, EXPLAIN TIPs are focused on events that occurred in the past, rather than future events—an explanation cannot be provided for something that has not happened yet. Interface designers must take this information into account when creating display components showing TIP status or progress. For instance, as ALERT TIPs are concerned with future states, they do not need to take up a portion of the display at all times in the same way a MONITOR TIP does.

TIPs States

Because TIPs drive user displays, it is necessary to define states and transitions to provide the interface with accurate and up-to-date information on a given TIP's state. For instance, after an explain TIP is called, how does the caller distinguish between that TIP being "called" versus "received and in progress" versus "complete and ready for human consumption". In the following subsections, we present state diagrams to cover different TIP types.

Alert, Monitor, and Request ALERT, MONITOR, and REQUEST share the same state diagrams due to the similar nature of their structure and outcome. Underlying both is a process whereby the automation assesses the value of some aspect of the world state—for instance fuel level, wind speed, or existence of adversarial entities—and reports it back to the human. In the case of ALERT, this is done when a predefined trigger condition is hit; in the case of MONITOR, this is done on a periodic schedule; and in the case of REQUEST, it is done once as soon as the human specifies. As with all TIPs, these start as `available`, when loaded from a TIP library during pre-mission planning, then either before or during the mission can be `staged`, where they are brought up by a user for manipulation pre-call, or `called`, where the user has sent the TIP out to callees for evaluation. During the mission, a called ALERT TIP goes through transitions between `active`, where the automation is actively evaluating the TIP to determine if the trigger conditions are met, at which point it moves through a `triggered` state and a `returned` state, where it waits until brought up in the user's display and is `presented` (see Figure 2). For MONITOR and REQUEST TIPs, there is no concept of being `triggered`, as MONITORs are sustained, periodic reports on information, and REQUESTs begin immediately after the human's call. Therefore, they move immediately from `active` to `returned` once the data has been gathered and sent to the human.

Explain EXPLAIN TIPs differ from ALERT, MONITOR, and REQUEST TIPs in that they cannot be satisfied by simply relaying a pre-specified information state in the world—instead, the human instructs the automation to explain some past decision point or time period, at which point the automation must perform some work to identify relevant

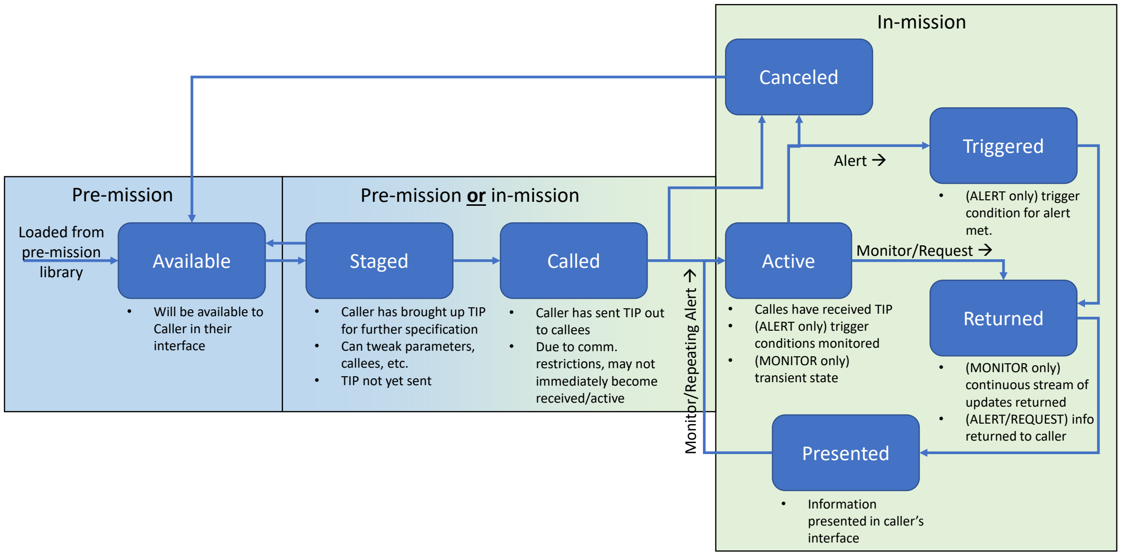

Figure 2: State diagram for ALERT, MONITOR, and REQUEST TIPs.

information and use it to generate that explanation. The automation sends the explanation to the human either on completion, or holds in a ready state to be returned to the human later on in the mission or during post-mission debrief. Therefore, the state diagram diverges from the previous TIPs after the `active` state into either `ready` or `failed`, depending on whether a valid explanation could be generated (see Figure 3). Note that although there is no `failed` state in the previous diagram (Figure 2), all TIPs can fail if, for example, communication between the human and automation does not allow for all the information to be transmitted in a timely manner. EXPLAIN TIPs simply have an extra, unique failure case that can occur even when communication is good.

Evaluate and Prepare EVALUATE and PREPARE TIPs ask the automation to assess the impact of some event, decision, or other piece of information on their current or planned execution. In the case of EVALUATE, this involves assessing whether the current plan is still valid or needs modification. For PREPARE, this involves generating a new plan that takes into account the provided TIP parameters. Although the automation's job may be complex, the lifecycle of EVALUATE and PREPARE TIPs are the simplest of the TIPs presented so far. Like other TIPs, EVALUATE and PREPARE can be `staged` and `called` in-mission or pre-mission, then once `active`, the automation works to either evaluate the impact to the provided information on their current plan or generate a new plan. The TIP moves to `ready` once the automation has completed its analysis or created the plan, and later the TIP information is `returned` and `presented` to the user (see Figure 4), much like an EXPLAIN TIP.

Example Scenario

We have developed an abstract domain, named "Lord of the Dragons." This domain has started as a variant of the Logistics (Veloso 1992) AI planning domain and focuses on adversarial decentralized planning and execution. In this scenario, three groups of dragons, overseen by elves, must carry rocks from their predesignated starting locations to a given goal location as specified in the initial state of planning problems. The initial state also describes geospatial constraints, such as boundaries of the area in which dragons work and terrain properties, which affect the solution plans and routes the dragons can fly to achieve their goals.

The domain can also be configured to include any number of adversarial agents, called orcs, that present obstacles for the plans of the dragons. Different classes of planning problems in this domain can be generated by varying positioning of the orcs, different capabilities that will harm the dragons, and the dragons' ability to neutralize them. there can be different breeds of orcs in a world and each breed is typically capable of using a suite of weapons such as spears and bow-and-arrows against dragons. Some orc breeds are capable to lure deafen dragons by the sweetness of their songs.

In our current version of the domain, we assume all of the orcs are stationary agents; i.e., they do not change their location from their initial positioning during a planning episode. However, a planning problem might be incomplete—an initial state is not guaranteed to specify the locations and capabilities of all the orcs that are present in the world.

As an example problem instance, consider the following scenario. On their ingress, the dragon team discovers unexpected pop-up orc units to the west of the known threat locations. They report this to their commanding elf, who activates a REQUEST TIP for more info about the orcs. This

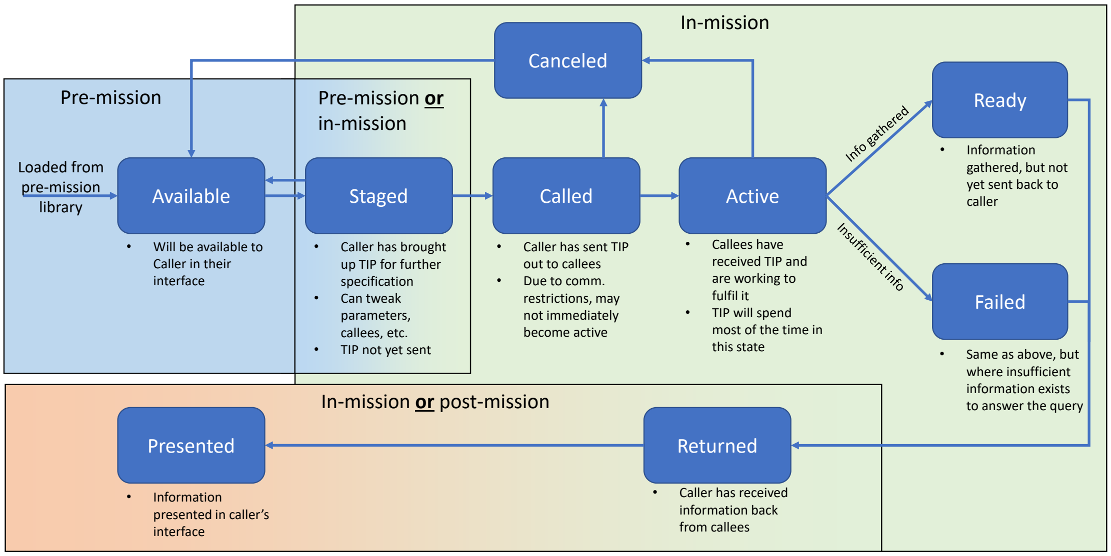

Figure 3: State diagram for the EXPLAIN TIP.

goes unanswered however, because the dragons were told to remain silent not to invoke the singing orcs, and this information was not conveyed to the elf in pre-mission planning. The elf asks for an EXPLAIN TIP to understand why the REQUEST TIP went unanswered. The answer may be provided immediately or deferred for review until after the flight.

The dragons change their formation so that the fire-breathing ones can clear the orc units and keep the other dragons safe. The commanding elf directs the dragons to an alternative flight route through the valley to the east of the world. TIPs provide the elf with notification of the new orcs and a pre-planned EVALUATE for plan variations (e.g., energy (food), time, and fire usage).

The forward dragon, A, in the flight pattern screams at the orcs to deafen them, whereas others, B and C, breath fire at the orcs. In the process, the elf anticipates energy and fire-breathing strength shortage and calls in a new flight of dragons: D, E, F, and leader G. Dragon C misses its target orc unit due to wind conditions. The lead dragon G engages. The commanding elf takes control of all of the the other dragons while dragon G is busy with the target. TIPs alert the elf about the energy drops on A, B, and C (due to the replan route). The elf directs these out of the area and gives the dragons D, E, F to G to control. The team neutralizes the orcs, arrives at their goal location, and drops their rocks.

We are planning to use this domain in our empirical and walkthrough analyses of TIPs and how helpful TIPs can be for a user. The following summarizes a "points of variance" example, describing what would happen with or without TIPs in a simple scenario as described above: on approach, the dragons monitor weather, their energy levels, location of other other dragons, and watch for threats. Monitoring time and attention would be much involved without TIPs, and could be theoretically be reduced with ALERT TIPs. Later in flight, after the orc units are discovered, Dragon C starts its fire-breathing maneuver. The overseeing elf inserts a TIP: ALERT if an orc unit is targeted and missed. Without this TIP, the elf has to monitor for effects continuously, significantly increasing their mental workload. The lead dragon destroys orc units and the elf gives the custody of dragons D, E, and F to the lead dragon. A new TIP triggers: ALERT if energy level is below 40%. In the absence of TIP, the elf may not remember to check for the energy levels until it is too late.

Conclusions

This paper presents the novel concept of Transparency Information Pacts (TIPs), which serve as contracts between human and automation under their oversight that aim to guarantee when and how certain information will be delivered to the human. The goal of TIPs is to provide a vehicle for improved transparency across the lifecycle of a mission.

To that end, we have described in detail how TIPs are structured, their requirements, and their output. We have identified several types of TIPs that could be implemented in human-machine teaming systems, although we anticipate future work will refine and expand this list as new types of information and new methods of conveying and achieving transparency are developed. Our scenario illustrates how TIPs could be used in a situation where one or more human operators are working alongside multiple autonomous vehicles, yet TIPs have wide applicability to any domain where humans and automation work in tandem, including industrial control systems, robot-assisted medical procedures, and cybersecurity.

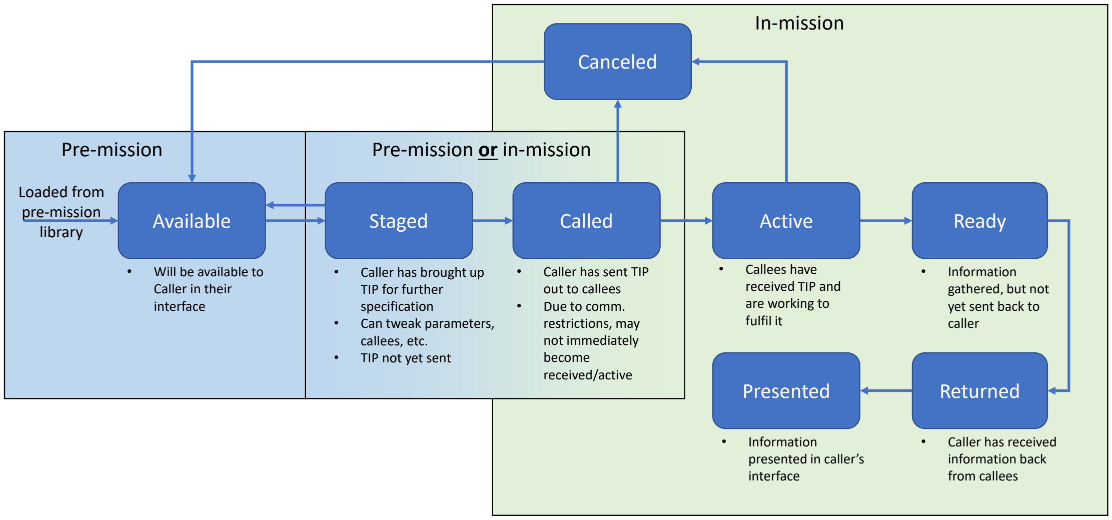

Figure 4: State diagram for the EVALUATE and PREPARE TIPs.

Future work will focus on implementing and integrating TIPs into existing simulations for human-machine teaming, with the eventual goal of deploying TIPs on real world systems. At the same time, we plan to continue research into improving lifecycle transparency, and expanding and refining TIPs as needed to account for new scenarios as they are uncovered. Hopefully, future human-machine systems will employ TIPs as an essential tool in improving coordination and situational awareness between humans and automation, thereby improving the capabilities of these systems and allowing both the human and automation to focus more time on the mission components to which they are each best suited.

Acknowledgements

This material is based upon work supported by the United States Air Force under Contract No. United States Air Force FA8650-22-C-6418. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Air Force.

References

Bhaskara, A.; Skinner, M.; and Loft, S. 2020. Agent transparency: A review of current theory and evidence. *IEEE Transactions on Human-Machine Systems*, 50(3): 215–224.

Chen, J. Y.; Lakhmani, S. G.; Stowers, K.; Selkowitz, A. R.; Wright, J. L.; and Barnes, M. 2018. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical issues in ergonomics science*, 19(3): 259–282.

Chen, J. Y.; Procci, K.; Boyce, M.; Wright, J. L.; Garcia, A.; and Barnes, M. 2014. Situation awareness-based agent transparency. Technical report, Army Research Lab Aberdeen Proving Ground.

de Visser, E. J.; Cohen, M.; Freedy, A.; and Parasuraman, R. 2014. A design methodology for trust cue calibration in cognitive agents. In *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments: 6th International Conference, VAMR 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I 6*, 251–262. Springer.

Ehsan, U.; Liao, Q. V.; Muller, M.; Riedl, M. O.; and Weisz, J. D. 2021. Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–19.

Entin, E. E.; and Serfaty, D. 1999. Adaptive team coordination. *Human factors*, 41(2): 312–325.

Fern, L.; and Shively, R. J. 2009. A comparison of varying levels of automation on the supervisory control of multiple UASs. *Proceedings of AUVSI's Unmanned Systems North America 2009*, 10–13.

Ho, N.; Sadler, G. G.; Hoffmann, L. C.; Zemlicka, K.; Lyons, J.; Fergueson, W.; Richardson, C.; Cacanindin, A.; Cals, S.; and Wilkins, M. 2017. A longitudinal field study of auto-GCAS acceptance and trust: First-year results and implications. *Journal of Cognitive Engineering and Decision Making*, 11(3): 239–251.

Kilgore, R.; and Voshell, M. 2014. Increasing the transparency of unmanned systems: Applications of ecological interface design. In *Virtual, Augmented and Mixed Reality. Applications of Virtual and Augmented Reality: 6th International Conference, VAMR 2014, Held as Part of HCI Inter-*

national 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part II 6, 378–389. Springer.

Lyons, J. B. 2013. Being transparent about transparency: A model for human-robot interaction. In *2013 AAAI Spring Symposium Series*.

Lyons, J. B.; and Guznov, S. Y. 2019. Individual differences in human–machine trust: A multi-study look at the perfect automation schema. *Theoretical Issues in Ergonomics Science*, 20(4): 440–458.

Lyons, J. B.; and Havig, P. R. 2014. Transparency in a human-machine context: Approaches for fostering shared awareness/intent. In *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments: 6th International Conference, VAMR 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I 6*, 181–190. Springer.

Lyons, J. B.; Koltai, K. S.; Ho, N. T.; Johnson, W. B.; Smith, D. E.; and Shively, R. J. 2016. Engineering trust in complex automated systems. *Ergonomics in Design*, 24(1): 13–17.

Lyons, J. B.; Wynne, K. T.; Mahoney, S.; and Roebke, M. A. 2019. Trust and human-machine teaming: A qualitative study. In *Artificial intelligence for the internet of everything*, 101–116. Elsevier.

Mercado, J. E.; Rupp, M. A.; Chen, J. Y.; Barnes, M. J.; Barber, D.; and Procci, K. 2016. Intelligent agent transparency in human–agent teaming for Multi-UxV management. *Human factors*, 58(3): 401–415.

Miller, C.; Goldman, R.; Funk, H.; Wu, P.; and Pate, B. 2004. A playbook approach to variable autonomy control: Application for control of multiple, heterogeneous unmanned air vehicles. In *Proceedings of FORUM 60, the Annual Meeting of the American Helicopter Society*, volume 3, 2146–2157. American Helicopter Society International, Inc Baltimore.

Miller, C. A. 2021. Trust, transparency, explanation, and planning: Why we need a lifecycle perspective on human-automation interaction. In *Trust in human-robot interaction*, 233–257. Elsevier.

Miller, C. A.; and Parasuraman, R. 2007. Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human factors*, 49(1): 57–75.

Ososky, S.; Sanders, T.; Jentsch, F.; Hancock, P.; and Chen, J. Y. 2014. Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. In *Unmanned systems technology XVI*, volume 9084, 112–123. SPIE.

Sadler, G.; Battiste, H.; Ho, N.; Hoffmann, L.; Johnson, W.; Shively, R.; Lyons, J.; and Smith, D. 2016. Effects of transparency on pilot trust and agreement in the autonomous constrained flight planner. In *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, 1–9. IEEE.

Shin, D. 2021. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, 146: 102551.

Simpson, A.; Brander, G.; and Portsdown, D. 1995. Seaworthy trust: Confidence in automated data fusion. *The Human-Electronic Crew: Can we Trust the Team*, 77–81.

Smith, C. J. 2019. Designing trustworthy ai: A human-machine teaming framework to guide development. *arXiv preprint arXiv:1910.03515*.

Stowers, K.; Kasdaglis, N.; Rupp, M.; Chen, J.; Barber, D.; and Barnes, M. 2017. Insights into human-agent teaming: Intelligent agent transparency and uncertainty. In *Advances in Human Factors in Robots and Unmanned Systems: Proceedings of the AHFE 2016 International Conference on Human Factors in Robots and Unmanned Systems, July 27-31, 2016, Walt Disney World®, Florida, USA*, 149–160. Springer.

Stowers, K.; Kasdaglis, N.; Rupp, M. A.; Newton, O. B.; Chen, J. Y.; and Barnes, M. J. 2020. The IMPACT of agent transparency on human performance. *IEEE Transactions on Human-Machine Systems*, 50(3): 245–253.

van der Vecht, B.; van Diggelen, J.; Peeters, M.; Barnhoorn, J.; and van der Waa, J. 2018. SAIL: A social artificial intelligence layer for human-machine teaming. In *Advances in Practical Applications of Agents, Multi-Agent Systems, and Complexity: The PAAMS Collection: 16th International Conference, PAAMS 2018, Toledo, Spain, June 20–22, 2018, Proceedings 16*, 262–274. Springer.

Veloso, M. M. 1992. Learning by Analogical Reasoning in General Problem Solving. PhD thesis CMU-CS-92-174, School of Computer Science, Carnegie Mellon University.