Displaced Interactions in Human-Automation Relationships: Transparency over Time

Christopher A. Miller(✉)

Smart Information Flow Technologies (SIFT),
319 First Avenue N., Suite 400, Minneapolis, MN 55401, USA
cmiller@sift.net

Abstract. Transparency (roughly, the provision of information about what the automated system is doing and why, potentially at multiple levels of abstraction and goal directedness) in automated systems has repeatedly been shown to improve human-machine interaction and performance, as well as human acceptance and trust. Nevertheless, there is a fundamental problem with providing transparency information at the time of action execution: specifically, that excessive human workload, which typically motivates the inclusion of automated systems, may not permit the absorption of the transparent information. We propose "displacing" the provision of transparent information in time and/or space from the time of execution and show how this approach is tied to beneficial findings for pre-mission planning and post-mission debriefing and explanations in human-automation interaction.

Keywords: Transparency · Explanation · Debriefing · Mission planning
Team interactions · Trust · Cognitive workload

1 Introduction

As the automated systems with which humans interact become more and more "autonomous" they, largely by definition, operate at times and in contexts in which human supervisor/controllers are not actively monitoring or issuing control inputs. This may be because inputs are impossible (such as in loss of communications due to jamming or weather or deep space communication time lags) or because humans are not able, willing or expected to intervene (e.g., due to workload constraints, inferior performance or simply to expectations of autonomous functioning).

These circumstances represent a separation of human input from execution, rather than a complete removal. Humans still design and build the systems and set them up (perhaps in a "mission planning" phase) to operate "autonomously" within certain constraints and for certain objectives. In the end, we humans still want the systems to behave for our benefit and within the instructions we provide.

Yet this displacement of control means that behavior shaping inputs and interactions will, increasingly, be displaced relative to action execution. The displacement may well be geographical—after all, one of the uses for unmanned systems is to have them operate in locations where it is dangerous for humans to go—but it will *inevitably* be temporal. We will be tasking automation increasingly via abstract guidelines, plans,

D. Harris (Ed.): EPCE 2018, LNAI 10906, pp. 191–203, 2018.
https://doi.org/10.1007/978-3-319-91122-9_17

and policies well before actions are taken and decisions are made, and we will be reviewing those actions and decisions (and, ideally, explanations for them) after they have been completed and their effects are known.

This means that what we have known about human-automation relationships must be re-examined for “displaced interactions”. We must ask questions such as the effect of varying degrees of temporal and geographic displacement on maintaining Situation Awareness (and what that means when communications are not possible or expected), what skills or personal traits are needed to accurately, effectively and safely interact with automated systems over increasing degrees of displacement, and what kinds of decision aids and training may make such interactions easier and more effective.

We have begun thinking about one such “displacement effect”—the role of transparency in displaced interactions. “Transparency”, here, refers to the ability for the automation to be inspectable or viewable in the sense that its mechanisms and rationale can be readily known. Researchers since at least Billings [1] and Sarter and Woods [2, 3] have called for greater transparency in automation. Furthermore, recent research has generally confirmed [4–7] that such transparency yields better human situation awareness, trust and, frequently, better overall human-machine performance than systems which include less or no transparency.

But there is a problem inherent in transparency that begs for an answer to the displacement question. As noted previously [8] not all information about what an “autonomous” agent is doing and why can be shared if there is to be any workload savings in a multi-agent team. Indeed, for the human to attend to and process any information about what an “autonomous” agent is doing will come at the expense of the human’s attention devoted to perceiving and understanding other aspects of their world and performing actions within it. We should expect this to produce additional workload and potential loss of situation awareness of other aspects of the work context if attention is oversubscribed.

Since automation is frequently introduced precisely to enable a human operator to do more within available time, workload (and training), the introduction of additional information for the human to process during this busy period may be particularly problematic. Worse, as noted above, there are contexts such as deep space exploration and military operations under communications jamming in which the communication of information to support automation transparency may simply not be possible.

So, we can conclude that transparency is valuable in human-automation interaction, but information to support it frequently cannot be communicated—at least in the moment when it is most needed–for a variety of reasons. Is there a way out of this dilemma?

2 Displaced Transparency

The problem represented by this dilemma is hardly new, nor unique to human-automation interactions. Humans interacting with other humans have encountered and wrestled with it throughout time. Human supervisors attempting to increase their capabilities through organizing and administering (human) subordinates have almost identical requirements: the need to maintain awareness of the performance of their subordinates so as to ensure that the supervisor's intent is accomplished as accurately as possible—even when that subordinate must necessarily act at a geographic or temporal distance from the supervisor.

As I have argued elsewhere [9], acting through a human subordinate is a process of delegation. Delegation inherently involves the communication of intent with oversight (including inspection and correction) of the subsequent performance of that intent by the subordinate. Furthermore, the communication of intent itself forms an "intentional frame" which can serve to increase situation awareness and decrease communication and cognitive processing demands in the future. Sheridan's original definition of supervisory control [10] included the provision that supervisors had to communicate their intent to subordinates or, in Sheridan's words, to "teach" subordinate automation what it should do.

Human to human communication in and near the moment of execution is an extraordinarily rich tool, especially when it is deployed against the backdrop of common cultural and professional understanding of the domain and its goals and methods, and even moreso when it is augmented by mutually-understood professional jargon. It serves to make the communication of intent from supervisor to subordinate, and status from subordinate to supervisor, extraordinarily efficient and rich. But it is worth noting that there is evidence that high-performing human-human teams, especially in high criticality domains, frequently exhibit less (and less explicit) communication than do less well-integrated teams [11].

Part of the reason human natural language communication can be so effective, and a large part of the reason well-trained and experienced teams can be so sparse with their communications, is that such team members share an understanding of the domain and of work within it. This understanding is certainly acquired during training and experience, but on a day by day (or mission by mission) basis, it is also acquired through mechanisms such as training and planning before execution, and explanation and debriefing (or after-action reviews) after execution. Below, I will argue that these mechanisms provide "transparent" information to the human team-members—that is, the same kind of information that has been shown to improve performance from "transparent" displays. They just do so at times other than the time of execution. That is, they provide temporally (and potentially, geographically) displaced information about how an agent is, will, or should behave, or how and why he/she/it did behave, but do so at a time when workload and attentional demands are lower (potentially on both supervisor and subordinate) than the time and context of execution. In short, they provide *Displaced Transparency*.

3 Transparency and Why Displaced Information Can Provide It

Chen et al. [12] have defined transparency as "…the descriptive quality of an interface pertaining to its abilities to afford an operator's comprehension about an intelligent agent's intent, performance, future plans, and reasoning process." But the emphasis on "the interface" in the above definition puts a, perhaps undue, focus on information that is conveyed during execution—when an interface is typically used. I contend that much information which achieves the goals of transparency (i.e., affording “an operator’s comprehension about an intelligent agent’s intent, performance, future plans and reasoning process”) need not be provided only, or even primarily, at that workload-intensive time.

Indeed, Chen [12] also defines a scale or model for transparency, the Situation Awareness-based Transparency (SAT) scale, which in turn leverages Endsley’s [13] scale for Situation Awareness. Chen’s SAT levels are summarized below as transparency information intended to support:

1. Level 1 SA (What’s going on and what the agent is trying to achieve) which is satisfied by providing information about the agent’s:
   - Purpose or Desire and Goal selection
   - Process and Intentions (including Planning and Execution) and Progress along that process
   - Performance of both the process and in general.
2. Level 2 SA (Why does the agent do what it does?) which is satisfied by providing information about the agent’s:
   - Reasoning process for planning or decision making, including the agent’s beliefs and broader purpose, including agent’s current beliefs about the environmental and other factors which constrain it
3. Level 3 SA (What should the operator expect to happen?) which is satisfied by information about the agent’s:
   - Projection to Future/End state
   - Potential Limitations including likelihood of error and history of performance.

If we take this information content as what is required for effective transparency, then it is worth noting that much of it could be—and in much effective human-human teaming, is—provided either before or after the time of action execution. “What an agent is trying to achieve?” is something that can and generally should be worked out, at least at a high level, before the subordinate agent is deployed. Intent expressions (which may include goals, purposes, methods and priorities [14]), by their nature, occur before action, while an understanding of why an agent does what it does and what its beliefs were that might have influenced decisions and actions are precisely the focus of the explanations that occur in effective after-action reviews and debriefings [15]. Figure 1 provides a hypothesized annotation of Chen’s SAT levels into items which can occur before or after the moment of action.

Figure 2 provides a simple timeline for a pre-planned workflow (e.g., for a military reconnaissance mission) to illustrate the point. If we assume mission execution to begin at time T, the let us say that at some point before T (i.e., time T minus n), a supervisor and subordinate plan this recon mission as consisting of an Ingress Phase to being at time T and end at time T+1:00 h, to be followed by a Search Phase to run from the end of the Ingress Phase for another hour. These will then be followed by a decision point whose agreed-upon logic is that if a target has been detected it will be Monitored for another hour and if not, the agent will Traverse to an egress point.

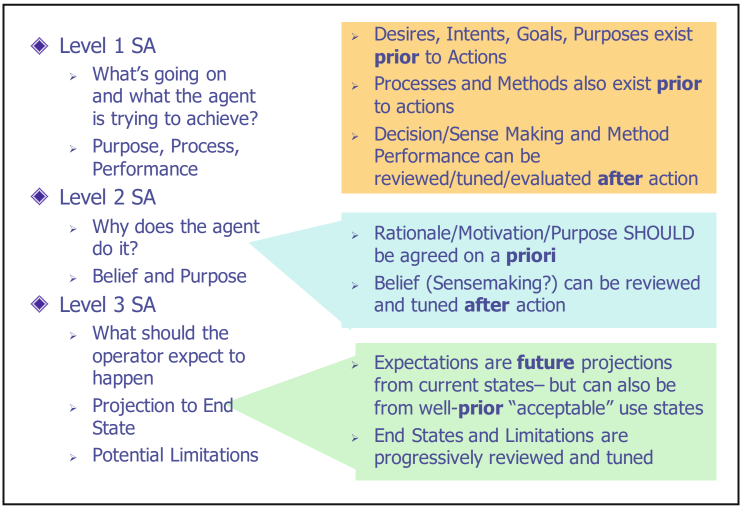

Fig. 1. Chen et al.'s [12] SAT levels annotated for their ability to be conveyed either before or after the time of action execution.

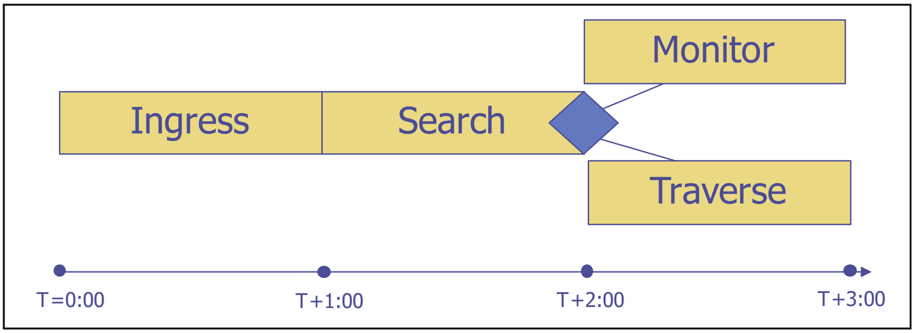

Fig. 2. Simple timeline example to illustrate conveyance of "transparent" information before use via planning.

The simple fact of having made this a prior plan affords substantial situation awareness to the supervisor *even if s/he has no further communication* with the subordinate. For example:

- Given that it is 45 min into the mission, the supervisor knows that the subordinate is engaged in Ingress and even, approximately, where the subordinate is. This is "What's going on?" knowledge—Level 1 SAT.
- Furthermore, the supervisor knows why the subordinate is ingressing: to get to the search area and begin search—Level 2 SAT.
- Finally, the supervisor knows that, at time 1:00, the subordinate will transition to Search. This is "What the operator should expect to happen?" knowledge—Level 3 SAT.

Granted, all of this SA can be in error if no further communication is available. Even the best laid plans can go awry, and lack of communication will make that deviation opaque (the opposite of transparent) to the supervisor. But we should also note that having an a priori plan in place makes communication more efficient. Instead of having to report all three levels of SA (what is going on, why, and what to expect next), the subordinate will generally have to report only current status or, perhaps, only deviations. The presence of pre-planned alternatives and acceptable error bounds (as represented partially by the pre-planned decision point in Fig. 2) makes it possible to further reduce communication needs. If, anywhere from time 1:00 to 2:00 the subordinate reports finding a target, then the behavior at the time 2:00 decision point becomes predictable. Even more, if at some time after 2:00 the subordinate reports that it is Monitoring, the superior is entitled to conclude that the reason is because a target was found during Search. Even in the event where the entire plan is made impracticable, having had a shared plan makes communication about future behaviors more efficient by giving all participants a shared ground to build from. If, for example, a fuel leak makes sustained Monitoring impossible, both supervisor and subordinate will know (at least approximately) where the agent is, how much range is required to get home, that a target has been detected, that the monitoring objective will have to be abandoned, etc. And all of this knowledge is shared with little or no "in the moment" information exchange or processing.

The above example emphasizes the role of pre-mission planning in establishing "displaced transparency", but that is not the only way transparent information can be displaced. After action reviews, debriefings and explanations also provide after-the-fact transparent information as well. Admittedly, this information is not provided in a timely fashion to enable a supervisor to override or correct behavior which may not be desired, but insofar as work with this particular subordinate continues in the future, it does play a mutual learning and trust building/tuning role. By learning how the subordinate thought and behaved in a specific situation, and potentially by offering advice or instruction about how future instances should be handled, the supervisor can shape future behavior in much the same way that planning shapes behavior (and information interpretation) before execution. As Tannenbaum and Cerasoli [15] have said, after-action debriefs are an effective and efficient means of improving team performance and team cohesion. Their meta-analysis of debriefing studies (covering 46 samples and 2,136 individual participants) indicated that on average, debriefs improve effectiveness over a control group by approximately 25% (d = .67).

Figure 3 illustrates the "displaced transparency" relationship we posit. Most transparency research to date focuses on information which is conveyed in a narrow temporal window around an action or event of interest. Such research has generally shown improvements in performance, situation awareness and trust when transparency information is provided. But such research has rarely examined workload effects, especially in time critical and overloaded periods and/or has examined it with subjective and coarse-grained tools such as NASA TLX [16]. We posit that there are periods in human-automation interaction, just as there are in human-human interaction, where the human's attention, processing and comprehension capabilities are so sparse and/or over-subscribed, that the attention to transparent information will, at best, be incomplete, and at worst may provide a disastrous distraction.

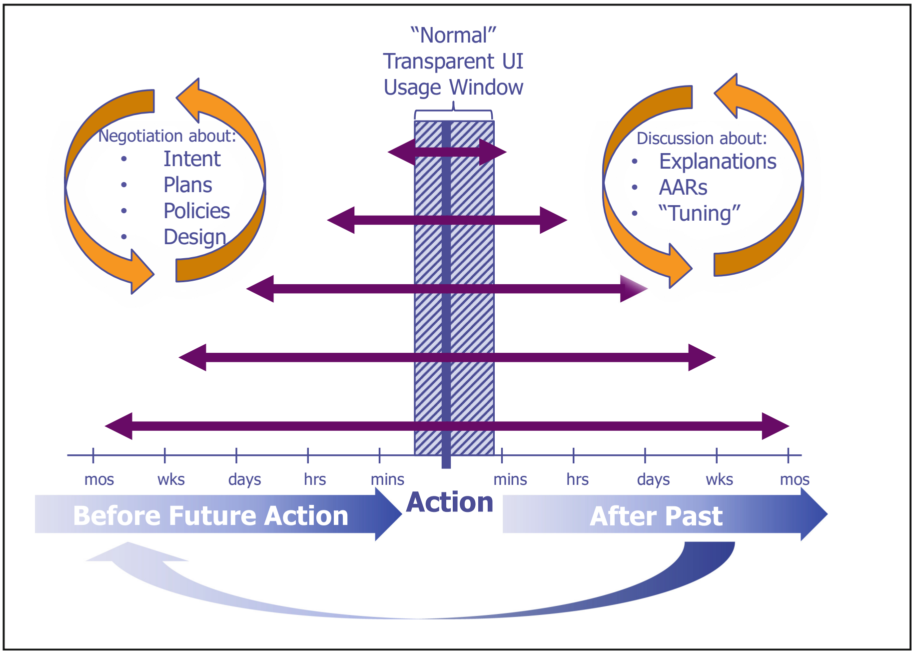

Fig. 3. Conceptual relationship between "in-action" transparent information and "displaced" transparent information.

We suspect, however, that it will be possible to displace much of this transparent information in time into periods which are substantially less overloaded. These periods may be anywhere from months or years before an action, or weeks to months after an action. The processes by which this transposition occurs are called training, discussion, planning, etc. when they occur before an action and the communication is about intent, plans, policies, alternatives and how to decide between them. At extreme durations, the "planning" process becomes one of design of the team, its concept of operations or its equipment. When the discussion occurs after the action, it is called explanation, debriefing, after-action review or various forms of "tuning" (including continuous quality improvement. Iterative cycles through either pre- or post-action discussion will likely only help the understanding of transparent information among team members. It is worth noting, in addition, that the process of pre-action planning and post-action review itself forms a virtuous cycle that can build team understanding for future actions. Although one might reasonably ask how and why providing transparent information *after* an action is helpful, the answer is that it will be helpful in providing team coordination and understanding for future actions, assuming that team must interact again in the future.

4 A Mechanism for Displaced Transparency and Its Effects

The claim that displacing transparent information can serve much the same purpose as presenting it at the time of need begs the question of how such a process might work. The work of Kambhampati, Chakraborti, Talamadupula and others [17, 18] may provide at least the beginnings of an answer.

Although working to provide explanations from machine or robot planners to humans, they nevertheless begin with a human social and cognitive model of the role of explanation itself. They take issue with many past approaches to providing machine explanations as rooted in presenting the machine's reasoning in its own terms—a process they call "soliloquy." [17] They say "Such soliloquy is wholly inadequate in most realistic scenarios where the humans have domain and task models that differ significantly from that used by the AI system." Instead, they say, effective explanation between actors must be a "Model Reconciliation Problem" [17].

Kambhampati has pointed our [19] that except perhaps in testing circumstances, explanations are generally neither sought nor provided in circumstances where individuals believe that their reasoning and behaviors are compatible or "reconciled". In other words, if I believe that I understand your mental model of the situation sufficiently that you and I would arrive at the same decision about a course of action to pursue, then no explanation of that decision is necessary. It is only when our models diverge (or are believed to diverge) that explanations are invoked—either because you do something which is not compatible with my model of what should be done in the situation, or because you intend to do something that you believe will differ from my model in the situation.

Explanation is thus, in their view, largely a process of Mental Model Reconciliation. It is provided in order to synchronize the models of those who must work together. It is sought when models do not synch—and it may not be sought when models are assumed to synch, even if they do not.

While explanations can certainly deviate from our actual methods of decision making [20, 21], they nevertheless represent how we are trained and acculturated to providing rationalizations for our decision making. Thus, it represents a form of team or group synchronization of thinking in its own right.

This viewpoint explains many phenomena in both explanation and transparency. For example, the fact that well-performing, efficient teams require less, not more, explicit communication [11] can be seen as arising from the fact that such teams are likely to have trained and worked together extensively in the past. Thus, their mental models of task and domain are likely to be well-synchronized—meaning that explanatory interactions are less likely to be required, commands can be abbreviated with contextual modifications presumed, and even task-based status information can be abbreviated and given with reduced contextual information because all parties are likely to understand what is needed when.

Similarly, team debriefing after a mission or project, with most of its associated benefits [15] can be attributed to the fact that effective debriefings involve team members interacting about their decisions and their behavioral processes. Simply knowing who knew what when and how they made decisions on the basis of that knowledge serves to educated other team members about their teammates' mental models, while group discussion about how things might be done more effectively in the future creates shared mental models going forward. Group discussion and shared rationalization of events and processes can also serve an educative function, effectively causing the group to converge on a series of broad thought processes and values that, within the "culture" formed by the group, count as valid and reasonable "ways of thinking" (or, at least, of explaining) [22]. Note, though, that this can cut both ways—resulting in adverse cases in "groupthink" [23] and "normalization of deviance" [24]. Finally, explanations offered within a group or to superiors play a social function as well [25], serving to reduce social distance and establish or reinforce power structures. While these functions may not directly contribute to a shared mental model about how decisions should be made within the group, they will serve to enhance team cohesion and affiliation.

On the pre-mission planning side, Dwight Eisenhower famously said "Plans are worthless, but planning is everything" [26]. One sense in which this is undoubtedly true is supported by the notion of model reconciliation. Extensively pouring over plans, rehearsing what could go wrong and what might be needed in contingent situations is a process which conveys and affords opportunities for model synchronization in much the same way that post-mission debriefing does. Participants are likely to emerge from such a process with a richer understanding of each other's models—and having had many opportunities to confront and refine those models in a process which tends toward synchronization. While this, too, can engender groupthink and the suppression of some voices, it also allows team members to "get inside each other's heads" and, thereby, increase their chances of knowing how each other will behave even in unanticipated situations.

Note that I am, in no way, claiming that measuring and computing model reconciliation needs and effects will be easy, especially between humans and machines. Chakraborti and Kambhampati and their colleagues have largely sidestepped this problem by using machine readable symbolic models for "simulated" humans in order to illustrate efficiency gains in symbolically characterized explanation content. By contrast, humans have evolved cultural, semantic and pragmatic markers rooted in natural language and "body language" for interpreting the need for model reconciliation and then effecting it—a process that, though complex and rich, is far from error proof (cf., [27] for examples of human-human and human-machine model mismatch where communication failed to avoid misinterpretations and, therefore, accidents).

5 Example of Efficiencies Through Model Reconciliation

It is reasonably straightforward to show how a process of model reconciliation which occurs either before or after a time frame in which the model is used to make a decision can lead to a reduction in the need to provide "transparency" information in that time frame. Consider again the simple mission sketched in Fig. 2, along with an "Observing Teammate" (OT) who, let's assume, can observe and know everything that is happening in the mission context but is completely unaware of the mission and the intentions of an "Enacting Teammate" (ET). In this scenario, any reported intentional interpretation of behavior from the ET will certainly increase OT's SA, but that is because OT's knowledge is essentially nil without such reports. Even observable events in the world (e.g., a headwind) will need to be interpreted for OT in order to provide Level 1 SA (Purpose, Process, Performance) knowledge, since the ability to interpret world events for their impact on the plan will be non-existent. OT has none of ET's mental model of the mission (though they may share a model of the world state in this example). All three levels of Chen's SAT [4] must be communicated for OT's full SA. On the other hand, if OT and ET share the same mental model and can perceive the same world events, then ET might need to communicate nothing since events would be interpreted similarly and decision making would be identical between ET and OT.

Even when OT can't observe everything that ET can (e.g., in the case of a remote supervisor), the burden of communication is substantially reduced. When events are unfolding as planned, at most ET might need to communicate confirmations that the plan is proceeding as expected. Even when unexpected events occur (e.g., the headwind above), reporting them may be all that is necessary to synchronize ET and OT's models of the impact and revisions necessary and desirable to the mission (revisions in knowledge at Level 3 SA that will *not* have to be communicated since both sides will make them concurrently—though confirmation might still represent useful redundancy). As before, anticipated variations (e.g., the decision point about whether to remain and monitor or simply to traverse) can be communicated much more tersely since both parties know, a priori, what the significance of a detected target will be on this decision point, and/or what the valid reasons are for remaining to monitor vs. traversing.

Even mental model mismatches become easier to detect given this prior planning. Let's say that OT failed to notice that ET detected a target and thus, doesn't understand why ET is transitioning to Monitor rather than Traverse. Simply posing the question in the context of what was expected to be a shared model of the mission identifies the mismatch. "Why are you Monitoring?" conveys a violation of expectations for Monitoring (i.e., the prior detection of a target) and hones in on the piece of information which is needed to repair model mismatch.

Finally, although harder to quantify, post-mission debriefings which are later followed by subsequent missions can have similar effects. If, for example, OT learns that ET has a tendency toward speedy completion of missions, s/he might assume a bias or preference in ET for Traversing vs. Monitoring and, in the presence of an ambiguous target detection signal, make more nearly accurate predictions about what ET will do. This represents a variation in the mental model (specifically, in values or priorities) between ET and OT, but insofar as OT understands this about ET (that is, OT's model of ET contains it) it will be accurately factored in to OT's SA and result in accurate understanding of the situation.

6 Predictions and Next Steps for Displaced Transparency

Above, I have presented an argument for the effects and desirability of displacing the presentation of transparency information into a priori mission planning interactions and a posteriori explanations and debriefings, along with a hypothesized mechanism for why these effects might be obtained. We know, from the sources cited above and others, that transparency information frequently provides detectable mission performance benefits. We also know (again from sources sited) that prior mission planning and explanation and debriefings also provide benefits for team cohesion, team satisfaction and team performance. It seems likely that these benefits obtain because they are making use of the same underlying mechanism: the communication of information which promotes situation awareness through mental model synchronization at the time of use. The fact that this information doesn't have to all be transmitted at the time of use, but instead can be spread into lower workload periods before and after usage, is a feature we should use more extensively in design—for human-human and also for human-machine interactions.

Some simple, testable predictions from this model are provided below. While we have not yet been able to conduct experiments to validate these predictions, a simple laboratory test seems eminently plausible. A relevant yet simple scenario is sketched in [18] where a human and a robot are located in a building with a long corridor and multiple side rooms. The human tasks the robot to fetch "a med kit"—one or more of which may be located in the side room(s). Which med kit is desired, expected, and provided is a function of elements of context (e.g., where robot, human, med kit(s) and other potential humans using med kits are located) as well as the robot's decision making algorithm. Of course, mismatches in elements of mental models (e.g., awareness of the physical context and of the robot's decision making process) are exactly what is required to provide SAT knowledge—and can be manipulated in an experimental design. The human may expect the robot to go to an different side room if s/he erroneously believes the med kit to be located there, or the robot may take longer and travel further than necessary if it has an erroneous model of where the human will be located. In this or a similar paradigm, we would predict:

- With mental model synchronization between teammates, reduced time, workload effort and even communication bandwidth will be necessary to achieve a similar level of situation awareness compared to conditions without mental model synchronization.
- Shifting the communication of transparent information into other time frames (before or after execution) will yield improved situation awareness (with reduced workload) even under conditions of communications restriction or constrained workload for the human recipient, given that model synchronization makes that information comprehensible.
- A priori mental model reconciliation will produce more accurate inferences by team members even in unanticipated situations, even with little or no explicit communication of transparent information.
- Particularly with regard to post-mission debriefing and explanations, effects of a posteriori model reconciliation will produce increased awareness and ability to predict teammates behavior even in unanticipated situations in subsequent missions.

We note with interest that [18] reports that the inclusion of mental modeling capabilities in the reasoning of a robot agent, where the robot was modeling the expected reasoning of a human operator and reacting accordingly, produced a 44–75% improvement in robot decision making in terms of avoiding resulting resource conflicts in one analytic experiment they performed.

What is less well documented is the tradeoffs involved in shifting transparency information into time frames before and after it is needed. Somewhere between "no plan survives first contact with the enemy" (implying that "overplanning" is wasteful) and "Plans are worthless but planning is everything" (implying that planning activities are very valuable), there must lie a (probably context-dependent) happy medium. Where is that medium, and what parameters characterize it? It is likely that information theoretic models can provide us with boundary conditions for this claim, but their relationship to actual human-human (or human-machine) interaction remains to be determined.

Finally, as automation becomes more capable, omnipresent and more "autonomous" in complex work domains, it is becoming clearer that it cannot provide all sufficient transparency information *in the moment* of action execution. Even if the human is capable of understanding it given time, it will all too frequently be the case that s/he will be engaged in other tasks and will be unable to devote sufficient attention and cognitive processing capability in a timely fashion. Instead, we need to strive to enable automation to participate in pre-mission planning and in post-mission debriefing and explanations in order to develop and accurately tune human trust and comprehension frameworks so that available capacity in the moment of use will be sufficient.

Acknowledgments. I am indebted to Dr. Jesse Chen for providing a forum for initial thoughts on this topic, and to Rao Kambhampati for the insight that explanations generally need to focus only on mismatches in mental models.

References

1. Billings, C.: Aviation Automation: The Search for a Human-Centered Approach. Erlbaum, Mahwah (1997)
2. Sarter, N.B., Woods, D.D.: How in the world did we ever get into that mode? Mode error and awareness in supervisory control. Hum. Factors **37**(1), 5–19 (1995)
3. Sarter, N.B., Woods, D.D., Billings, C.E.: Automation surprises. Handb. Hum. Factors Ergon. **2**, 1926–1943 (1997)
4. Mercado, J.E., Rupp, M.A., Chen, J.Y., Barnes, M.J., Barber, D., Procci, K.: Intelligent agent transparency in human–agent teaming for Multi-UxV management. Hum. Factors **58**(3), 401–415 (2016)
5. Lyons, J.B., Havig, P.R.: Transparency in a human-machine context: approaches for fostering shared awareness/intent. In: Shumaker, R., Lackey, S. (eds.) VAMR 2014. LNCS, vol. 8525, pp. 181–190. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07458-0_18
6. Ososky, S., Sanders, T., Jentsch, F., Hancock, P., Chen, J.Y.: Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. In: SPIE Defense + Security, p. 90840E. ISOP (2014)
7. de Visser, E.J., Cohen, M., Freedy, A., Parasuraman, R.: A design methodology for trust cue calibration in cognitive agents. In: Shumaker, R., Lackey, S. (eds.) VAMR 2014. LNCS, vol. 8525, pp. 251–262. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07458-0_24
8. Miller, C.A.: Delegation and transparency: coordinating interactions so information exchange is no surprise. In: Shumaker, R., Lackey, S. (eds.) VAMR 2014. LNCS, vol. 8525, pp. 191–202. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07458-0_19
9. Miller, C., Parasuraman, R.: Designing for flexible interaction between humans and automation. Hum. Factors **49**(1), 57–75 (2007)
10. Sheridan, T.: Supervisory control. In: Salvendy, G. (ed.) Handbook of Human Factors, pp. 1244–1268. Wiley, New York (1987)
11. Entin, E., Serfaty, D.: Adaptive team coordination. Hum. Factors **41**, 312–325 (1999)

12. Chen, J.Y., Procci, K., Boyce, M., Wright, J., Garcia, A., Barnes, M.: Situation awareness-based agent transparency (No. ARL-TR-6905). ARL/HRED Aberdeen Proving Ground, MD (2014)
13. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum. Factors **37**(1), 32–64 (1995)
14. Miller, C.: Delegation for single pilot operation. In: 2014 HCI-Aero. ACM, New York (2014)
15. Tannenbaum, S.I., Cerasoli, C.P.: Do team and individual debriefs enhance performance? A meta-analysis. Hum. Factors **55**(1), 231–245 (2013)
16. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Advances in Psychology, vol. 52, pp. 139–183. North-Holland (1988)
17. Chakraborti, T., Sreedharan, S., Zhang, Y., Kambhampati, S.: Plan explanations as model reconciliation: moving beyond explanation as soliloquy. In: Proceedings of IJCAI, pp. 156–163 (2017)
18. Talamadupula, K., Briggs, G., Chakraborti, T., Scheutz, M., Kambhampati, S.: Coordination in human-robot teams using mental modeling and plan recognition. In: IROS 2014, pp. 2957–2962. IEEE (2014)
19. Kambhampati, S.: Personal communication, Arlington, VA, 1 August 2017
20. Tversky, A., Kahneman, D.: Judgment under uncertainty: heuristics and biases. Science **185** (4157), 1124–1131 (1974)
21. Klein, G.: Naturalistic decision making. Hum. Factors **50**(3), 456–460 (2008)
22. Miller, C.: Learning to disagree: argumentative reasoning skill in development. Ph.D. thesis. University of Chicago, August 1991
23. Turner, M.E., Pratkanis, A.R.: Twenty-five years of groupthink theory and research: lessons from the evaluation of a theory. Organ. Behav. Hum. Decis. Process. **73**(2–3), 105–115 (1998)
24. Vaughan, D.: The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA. University of Chicago Press, Chicago (1997)
25. Brown, P., Levinson, S.C.: Politeness: Some Universals in Language Usage, vol. 4. Cambridge University Press, Cambridge (1987)
26. Eisenhower, D.D.: From a Speech to the National Defense Executive Reserve Conference in Washington, D.C., 14 November 1957. Eisenhower, D.D.: Public Papers of the Presidents of the United States, p. 818. National Archives and Records Service, Government Printing Office (1957). https://en.wikiquote.org/wiki/Dwight_D._Eisenhower. Accessed 3 Mar 2018
27. Miller, C.: Social relationships and etiquette with technical systems. In: Withworth, B., de Moor, A. (eds.) Handbook of Research on Socio-Technical Design and Social Networking Systems, Information Science Reference, Hershey, PA, pp. 472–486 (2009)