Communicating Unnamable Risks:

Aligning Open World Situation Models Using Strategies From Creative Writing

Beth Cardier
Griffith University / Trusted Autonomous Systems DCRC, QLD Australia
Eastern Virginia Medical School, Norfolk, VA, USA
b.cardier@griffith.edu.au, bethcardier@hotmail.com

Abstract

How can a machine warn its human collaborator about an unexpected risk if the machine does not possess the explicit language required to name it? This research transfers techniques from creative writing into a conversational format that could enable a machine to convey a novel, open-world threat. Professional writers specialize in communicating unexpected conditions with inadequate language, using overlapping contextual and analogical inferences to adjust a reader's situation model. This paper explores how a similar approach could be used in conversation by a machine to adapt its human collaborator's situation model to include unexpected information. This method is necessarily bi-directional, as the process of refining unexpected meaning requires each side to check in with each other and incrementally adjust. A proposed method and example is presented, set five years hence, to envisage a new kind of capability in human-machine interaction. A near-term goal is to develop foundations for autonomous communication that can adapt across heterogeneous contexts, especially when a trusted outcome is critical. A larger goal is to make visible the level of communication above explicit communication, where language is collaboratively adapted.

Introduction

Language shows itself, indeed, as the gateway to the mystery of the unsayable beyond language.
- Franke, 2014, 64.

How can a machine warn its human collaborator about a risk that it cannot explicitly name? This paper develops a conversational method based on three creative writing strategies to enable a machine to communicate an unexpected risk. During a mission, humans and machines can align their understanding of a situation using communication – an exchange of information tokens (Kennedy and Hidalgo 2021). However, open world situations are more complex than a machine's reference ontology can represent, and the simple vocabulary and dialogue currently supported by artificial intelligence might not be enough to communicate a problem that was unanticipated (Kruijff et al. 2015). To address this, a communication approach is designed for the open world that assumes the machine's underlying ontology will be insufficient and must be adjusted on the fly. In this adaptation, a shared situation model is collaboratively modified using conversation. A feature of this approach is its bi-directionality: multiple rounds of clarification are required between human and machine to achieve an unexpected adjustment of the shared situation model. These adaptive mechanisms are drawn from a domain that has not yet been leveraged for this problem: creative writing, as it is expressed in narrative.

Creative writing routinely conveys unexpected conditions (Herman 2002, p. 85) and does so without adequate language tokens. Booker-prize winning writer, Salman Rushdie, observes that "A poet's job [is] to name the unnamable" (Rushdie 2008, p. 3). Author Milan Kundera notes that good writing captures "something which hasn't already been said, demonstrated, seen" (Elgrably and Kundera 1987, p. 4). Creative writers have developed strategies to communicate novel ideas, using sophisticated inference structures to adjust a reader's situation model to include them. Three kinds of adjustment are explored in this method, one based on context and two based on analogy. The rationale for these strategies will be described theoretically and then consolidated into a conversational method.

It is not currently possible for artificial intelligence to communicate a condition that it has not been given the words to describe, whether embodied in a robot or not (Kruijff et al. 2015). Machines lack the ability to adapt to unexpected events because they depend on explicit terms and static frameworks (Sowa 2010) and their goals are limited by the way their capabilities are coded in advance (Johnson et al. 2018). In addition, machine communication

cannot yet emulate natural back-and-forth interaction (Li et al. 2016), making it difficult to maintain a common understanding as a situation changes. In spite of these limits, ongoing 'human-in-the-loop' communication with machines is critical to maintaining trusted performance (Kennedy and Hidalgo 2021). The need for adaptive capabilities have been identified as critical by the National Security Commission on AI 2021, which forecasts that by 2025, human-machine teams should be able to "reason over changing contexts to flexibly adapt teaming strategy" (Schmidt et al. 2021, p. 85).

This method accepts the limits of a machine's reference framework as a starting point. It assumes there will always be something missing from that reference framework when the machine operates in an open world, because the problem of inadequate language is a permanent feature of the open world. Expert human communicators accept this as part of our "common human experience of butting up against the limits of language" (Franke 2014, p. 23). To engage with this issue, this approach presents a scenario that occurs five years in the future, using engineering methods that could come within reach during that time, to imagine how current capabilities might be extended.

Example: A Ruined City

In this future scenario, an explosion has destroyed the edge of a metropolitan city and spread toxic chemicals into the air. In the hour immediately afterwards, when the need to find survivors is most urgent, it is not safe for human rescue teams to enter the area en masse. Fortunately, each emergency response center in this city has autonomous drones and robots that can locate possible survivors, so a minimal human team can follow with a targeted extraction.

One of these robots reports back to its human collaborator with an unexpected problem. The robot has located a human survivor lying beneath bricks and it tells its collaborator this. That discovery is an expected goal, so the machine can easily refer to it using explicit language tokens such as "survivor", "bricks" and "trapped".

However, the machine does not have the vocabulary to communicate an unexpected aspect of the situation, one that indicates urgent risk. A water pipe broke during the explosion and due to an accident of architecture, where a cement wall encircles the area, there is water pooling around the survivor, who is unconscious. Assuming the robot can visually identify and reason about some of the physical features surrounding the survivor, how can it convey this unexpected threat to its human controller?

To address this, the proposed method transfers adaptive structures from creative writing into a human-machine communication strategy. This approach engages with the gaps in the machine's *ontology,* just as a writer does for their human reader. In a machine, an *ontology* is the reference framework from which the machine operates, which functions as a kind of 'dictionary'. It is represented as "entities that are assumed to exist in [a] domain of interest as well as the relationships that hold between them" (Gruber 1995, p. 908).

In this scenario, the machine has reached the limit of its ontology because it has encountered an unexpected risk. This method *uses the edge* of that ontology to communicate. It does this by indicating the ways in which qualities of the unexpected condition fall inside or outside that limit. Conceptually similar work can be found in (Goel, Fitzgerald, and Parashar 2020), in which numerous analogical mappings are 'pruned' until a correct mapping is found. Except in this case, the pruning predominantly occurs in the human's mind, while the 'hints' that enable pruning are supplied by the machine.

During that pruning process, the human guides the machine to provide increasingly relevant hints, by focusing on the provided details around the risk condition. The human thus uses their powerful inference abilities to infer how to assemble the information they are given and formulate questions in response to gaps in the shared situation model.

The Conversational Process

The process begins when the machine **flags** an unexpected condition. The next phase uses three creative writing strategies to enable the human to understand what aspect of the situation contains the unexpected risk. These three strategies are: **context** (in which information from different contexts is supplied), **nearest partial match** (in which the ontological terms which were considered but rejected are listed) and **analogy** (in which the reason why ontological features matched and did not match the unexpected condition are explained). These strategies may repeat in any order. When the unexpected condition is adequately identified, the human extends the machine's ontology by **naming** the unexpected condition, combining known information tokens to create an ad hoc term to refer to it. The human also issues instructions or information to go with this new term. To confirm the new shared understanding, the machine **offers** to modify a shared plan using the new term and its own vocabulary. The human then **tests** the machine, trying out the new term in different contexts until they reach a condition for which the machine determines the new term does not apply. When the human agrees with these proposals – or not – they indicate the success of the alignment and the amount of trust now possible.

Below is an example interaction, set in the above scenario of a ruined city.

Machine: Our goal was to locate human survivors. I have located one but there is an unexpected condition. In combination, its features suggest risk. Would you like to hear? (*Flag)*
Controller: Proceed.

Machine: A man in a white shirt has been found lying beneath bricks. He is located in a building that has a curved wall but does not have a roof. (*Context)* The missing roof is unexpected. There is water in the situation. The water is unexpected. (*Flag)*
Controller: Describe the water.
Machine: The water has a partial match with the following terms: rain puddle, swimming pool, tide and flood. However, it is not recognised as any of these. (*Near match)*
Controller: Out of those options, which is the most similar to the water in this situation?
Machine: The water is most similar to a flood and a tide. (*Near match)*
Controller: Why did you reject these terms?
Machine: The water is rising like a flood but its area is less than two acres and so it is too small to be a flood. The water is spreading like a tide, but it is not connected to an ocean or river, and so it is not a tide. (*Analogy)*
Controller: What entities are close to the water? (*Context)*
Machine: Close to the water are: a human, a backpack, a curved wall, a pipe.
Controller: Is the water touching the pipe? (*Context)*
Machine: Yes.
Controller: We will define this water as 'pipe water' *(Naming).* Is the pipe water touching the human in the white shirt? (*Context)*
Machine: Yes.
Controller: Is the human awake? (*Context)*
Machine: No.
Controller: The pipe water might be a risk to the human. Prioritize the rescue of this survivor because they might drown *(Naming).* Confirm please (*Test).*
Machine: I will notify the rescue team that the 'pipe water' might be dangerous to the survivor, and that they should prioritize that rescue. Is this correct? (*Offer)*
Controller: Yes.
Machine: Notification of command complete. Rescue has been alerted.
Controller: Going forward, when you detect 'pipe water' you should follow the procedures for 'water'. (*Naming)*
Machine: Confirmed.
Controller: Let's test the new word 'pipe water'. Would you touch the pipe water? (*Test)*
Machine: Only if unavoidable. I will follow my procedure for any water, in terms of touching the pipe water. *(Offer)*
Controller: Correct. Would you consider pipe water to be a risk if electrical wires are present? (*Test)*
Machine: Yes. I will follow my procedure for any water, in terms of the pipe water's proximity to other risk conditions, like electricity. *(Offer)*
Controller: Correct. If I was standing next to pipe water, would you consider me to be at risk? (*Test)*
Machine: I will follow my procedure for any water, in terms of you standing next to the pipe water. I would report the situation to you in case you were unaware of the pipe water and wait for further guidance from you. (*Offer)*
Controller: Correct. That is all.

In this exchange, the human guides which information is supplied by the machine, by asking questions around the unexpected condition. The machine supplies details according to creative writing strategies that can adapt the human's situation model: *context*, *near match* and *analogy*. The human cycles through them, gathering different kinds of information until they can make a judgement about potential risk. If required, the human can extend the machine's ontology by adding a new term (in this case, 'pipe water'). Finally, that alteration is tested through a series of questions, to find its limit.

The processes of inference by which this occurs will now be explained in more detail, by finding aspects of the phenomenon in existing literature in domains related to cognition and communication, as part of a literature review.

A Bridge Through the Literature

Professional writers face a problem similar to that of the autonomous machine in unfamiliar terrain: they want to communicate information that is not shared by their audience and can only be conveyed using inadequate language tokens. To name the unnamable, creative writers incrementally provoke transformations in the reader's situation model, altering its network of knowledge structures to form new relationships. In narrative, this adjustment occurs in relation to unexpected information (Graesser and Wiemer-Hastings 1999, p. 78).

In psychology, the starting point for this transformation is the human's **situation model**, which is a "representation of relations of interest in the world" (Lambert et al. 2008, p. 1) that an individual generates to reason about aspects of a particular experience. It is constrained by the mechanics of perception, and in this sense, has features of context, in which a "limited part of reality" (Devlin 2009, p. 238) constrains a "deduction that is justifiable under one set of circumstances may be flat wrong in a different situation" (Devlin 2009, p. 2). Maintaining an accurate situation model is critical for appropriate decision-making, as a decision that is desirable in one situation might be harmful in another. Maintaining a shared situation model with a co-worker is necessary for trusted collaboration, as each must operate from a common reference to make co-operation possible.

A human's situation model is adjusted using inference. This requires a shift into the domain of psychology, which explains that inferences can create a situation model (Zwaan et al. 1995), integrate information into it (O'Brien and Cook 2016), bridge gaps presented by it (Asher and Lascarides 1998; Irmer 2011), indicate reference points for interpretation (Kahneman and Tversky 1979) and indicate how people

use existing tokens to imply implicit conditions (Elder and Haugh 2018). Out of these different forms of inference, the **bridging inference** is the focus here, due to its role in supplementing implicit details among explicit tokens.

A bridging inference is a mental projection that fills the gap from one statement to the next, enabling a person to relate entities or events "in a particular way that isn't explicitly stated" (Asher and Lascarides, 1998, p. 83) to make the text coherent (Clark 1977). Bridging inferences play a crucial role in enabling a human to connect fragments of information (Irmer 2011) especially if explicit connections are missing, as in the case of an author communicating a novel concept. Inferences can be direct or indirect, with indirect being more important to the process described here. In an indirect inference, referents are not explicit, so the implicit associations surrounding statements must be leveraged instead ("I've just arrived. The camel is outside and needs some water" (Asher and Lascarides 1998)). Indirect bridging inferences are the means by which a human's situation model can be adjusted beyond its existing scope, even if many details are not explicitly stated.

In this method, indirect bridging inferences are activated using storytelling devices - one related to **context** and two based on **analogy**.

To communicate any situation, the writer provides information from different **contexts** to describe it. These contexts can be physical, conceptual, historical or other kinds of information. Psychology has examined this use of numerous sources to build a situation model, and the way it which can consist of any kind of information, from "concrete geospatial relations through to abstract political relations" (Lambert et al. 2008, p. 1). Discourse analysis explains how "each piece of incoming information can be mapped onto a developing structure to augment it" (Gernsbacher 1996, p. 4). This enables a person to monitor "multiple dimensions of the evolving situation" (Zwaan et. al 1995, p. 395) so that contexts with different properties can "mutually constrain each other" and disambiguate interpretation of an situation, making comprehension more "fluent" (Zwaan 2016, p. 1030). More research is needed to understand how each bridging inference can act as a stepping-stone towards an unexpected condition.

By itself, a collection of statements will not naturally lead to the comprehension of novelty. The right inferences must be activated, one after another, to transform a person's situation model *in a specific direction*. In creative writing, this process is directed by another device – the strategic introduction of anomalies.

Narrative is predicated on a divergence from routine. This has been recognised in fields such as linguistic psychology, discourse processes and cognitive narratology. A writer directs their reader's bridging inferences by introducing anomalies. Anomaly establishes a background context (the **expected** status quo) and an agent in the foreground who takes the **unexpected** path, recognizable for the way it diverges from the regular frame (Herman 2002, p. 90). In cognitive science, this relationship is referred to as landmark (context) and trajector (agent) (Langacker 1987), where the landmark provides a point of reference "for locating the trajector" (Langacker 1987, p. 1:217). The story follows the agent through this divergent state, providing details at every step, so the reader can understand both agent and situation, and focus their inference on the details outside normality.

A different device, **analogy**, can assist with communication of this anomaly. The way this operates is complex. Analogy communicates "by a process of structural completion: learners transfer information from a more complete source domain to a target domain missing that information" (Clement and Yanowitz 2003, p. 196). It is a structure that enables the transmission of information by bridging two sources (Fauconnier and Turner 2002). This combination of known affordances with new effects enables the reader to derive an understanding of what kind of entity might have caused that change, even if it cannot be explicitly named. There is thus a combination of analogical mapping and new direction required to communicate novelty, in a manner that has been recognized in (Goel et al. 2020, p. 24). Here, adaption "from the known case to the new problem' is supplemented by "strategies of social learning" where a human may teach the robot a new activity using known structures. Analogy has already been explored in human-machine communication, with an extensive survey provided by French (2002). However, in these systems, the focus is on structure-matching rather than the communication of novelty.

Alignment among situation models is the ultimate goal of this approach, as it is an important enabler of trust. A common definition of trust is "the attitude that an agent will help an individual's goals in a situation characterized by uncertainty and vulnerability". For human-machine teaming, this definition is extended with open world autonomous systems in mind. This paper defines *trust* as resulting from a transfer of relevant information among the aligned situation models of collaborators, such that the degree and fidelity of transfer among them is adequate and verifiable for the required outcomes, in a particular set of conditions (context), *which are characterized by uncertainty and vulnerability*.

Past studies of trust have focused on other criteria, such as reliability (Hancock et al. 2011), transparency (Baker et al. 2018) and physical appearance (Song and Luximon 2020). A more complex conceptualization exists in literature on *rapport* in human-machine teams, which examines the transfer of agency that becomes possible due to an alignment of purpose (Bronstein et al. 2012). This approach better suits the problem of trust as a product of communication, which is the concern here.

Gibson et. al note that such cooperation is the product of multiple points of engagement, where one person is "orientated to" the terms of a relationship if it agrees with their

priorities and the task (2017, p. 306). Communication establishes these channels of affordance by registering common points of interest and/or values between the two parties. The alignment of trust forms in tandem with the emerging shared situation model, which unfolds as collaborators successfully exchange information.

Within this process of checking the alignment between situation models, a device from narrative improvisation is used: the 'offer' (Johnstone 1981). Acting improvisation theorist Keith Johnstone's developed the term 'offer' to refer to any "initiating move" such as a "physical or vocal action, which presents more information to a partner" (Garrett 2007, p. 9). In this paper's method, the machine makes an offer to perform a helpful action after the human has named the unexpected condition. The purpose is twofold: the proposed action will support the human, but in addition, enables the human to check whether the newly derived language token ("pipe water") has been correctly registered by the machine before it performs any urgent actions.

There are two different processes of alignment in this method. The first occurs after the machine flags the unexpected condition and the human questions it to discover how to adjust their situation model. The second occurs during the test stage, when the human checks whether the robot has correctly registered the newly named term. The goal is to find the limit of the application of the new term, where the robot would refer back to the human for further instructions.

The following section demonstrates how bridging inferences incrementally adjust the human's situation model around the unexpected condition.

Stages of the Example Conversation

The machine commences by **flagging** that there is an unexpected condition which suggests risk. This primes the human's ability to combine information from numerous distributed sources to fill in the gap. See Fig. 1 for the first two stages of this process. Stage 1 depicts the human's inferred situation model, which forms in response to that flag. The outer disk represents the shared situation model of the human and machine, and as such, is the situation from which subsequent expected inferences are drawn. Within the shared situation model, a circular zone indicates the flagged possible risk (the dotted line indicates there is inadequate information about it). The background context of the ruined city informs subsequent information inferred by the human.

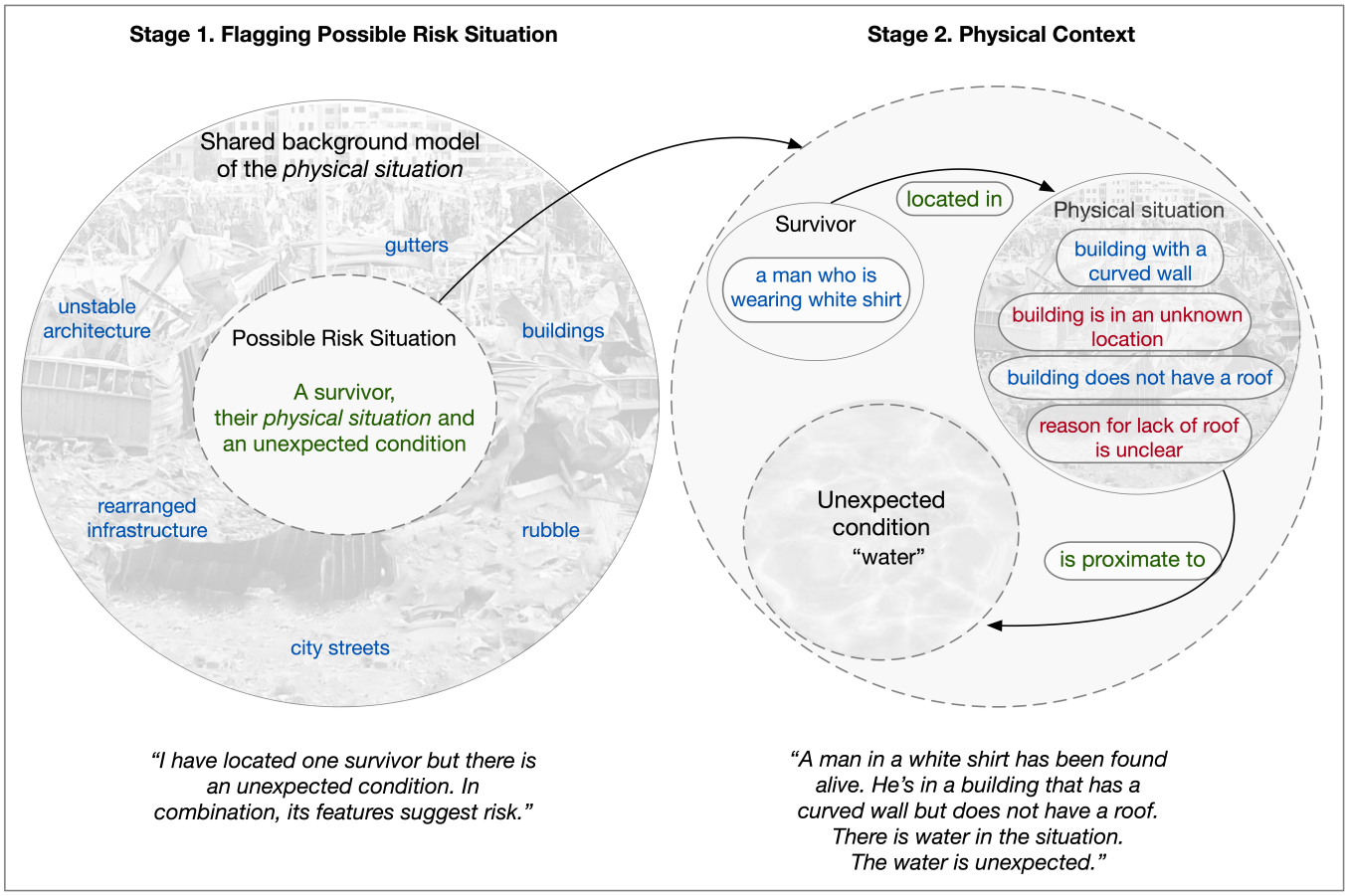

Figure 1: The first two stages, in which a machine communicates an unknown condition that suggests urgent risk.

In Stage 2, after the human confirms that they want to hear about the unexpected condition, the machine lists entities that are **expected**, to provide a physical context for any ambiguous information that will follow. These physical entities are oriented around the goal – in this case, the human survivor. The machine selects those tokens according to overlapping criteria, which enables it to prioritize entities for mention. For a rescue scenario, the suggested criteria are:

1) **Proximity**: the reported entity is physically or conceptually proximate to a goal, such as a survivor or a risk condition. Two levels of physical proximity are featured in this example: the immediate surrounding items ("lying under bricks"), and a larger scale ("He is located in a building which has a curved wall but does not have a roof").

2) **Goal**: the reported entity is related to the goal.

3) **Risk**: the reported entity has potential for intrinsic threat to human physiology (eg. smoke, water, fire, electricity, height, suspended weight).

4) **Rarity**: the reported entity appears less often in the context. Initially, 'expectedness' will be a predetermined set of terms. If those entities are also ranked according to likely appearance, those which are a lower likelihood in that context can be prioritized for mention.

These categories interact to prioritize tokens for mention. If an item is near the goal (human survivor) and also qualifies as a potential risk (water), it will be mentioned. By contrast, if the item is low risk and far away from the survivor, it is unlikely to be mentioned. A particular entity can thus be mentioned differently according to its role in the scene.

An adjustment to the human's situation model occurs when two risk entities "water" and "missing roof" are reported by the machine as being unexpected (Stage 2). Expected background information recorded in the Stage 1 situation model, such as "rearranged infrastructure," carries over to Stage 2 unless explicitly contradicted. This carryover enables the human to assume the missing roof reported by the machine might be due to the explosion, and so that line of questioning is not pursued (but could be). Instead, the human focuses on the more anomalous detail: the large volume of water on a city street, which potentially carries risk. The water thus becomes the focus of the human's questions.

The human subsequently requests more information about the water ("Describe the water").

Fig. 2 depicts the three further adjustments to the human's situation model (Stages 3-5). Stage 3 shows how this situation model becomes more detailed when the machine provides the list of words with which the water partially matched. These are near-match tokens for the unexpected condition of water that were considered but rejected: rain puddle, swimming pool, bathtub, tide, flood. Each item on this list characterises the near edge of the unexpected condition through a partial match.

This strategy provides an implicit scope of the qualities of the unexpected condition, provoking the human's bridging inferences to speculate about what parameters these

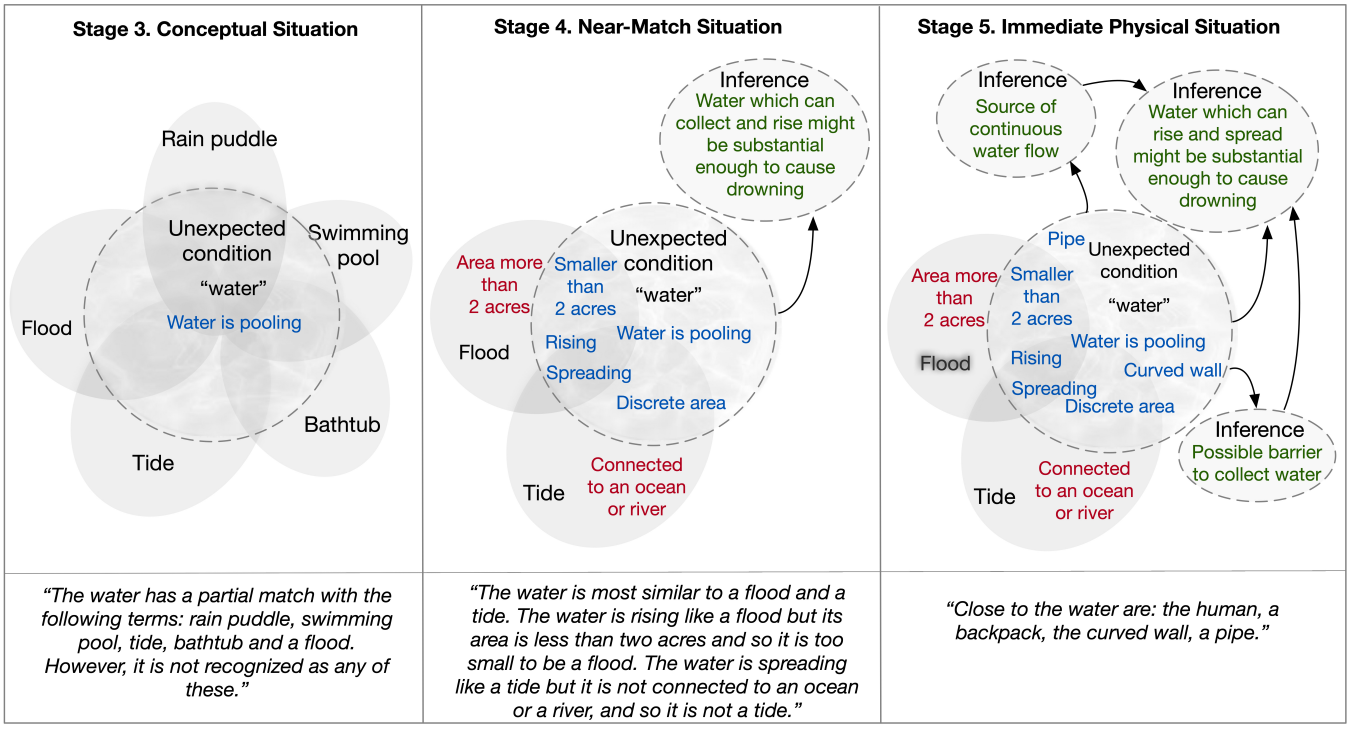

Figure 2: Three further stages, where enough information about the unknown risk is communicated to support a decision.

terms might have in common. In implementation, these near-match terms could be generated with current algorithms – vision-based classifiers using deep neural networks will generate the percentage match with known labels, and the highest matches could be provided.

The human responds by asking which of these terms were the *closest* match to the unexpected condition. Stage 4 shows the effect when the machine chooses two items with the highest match of features ('the water is most similar to a flood and a tide'). The human's situation model narrows further to the overlapping features for the terms 'flood' and 'tide'. Qualities of pooling and spreading are inferred. These new qualities increase the likelihood of the risk of the survivor drowning without yet confirming it. More information is needed.

The human then asks why these terms were rejected. The response in Stage 4 (lower left) leverages the human's ability to reason using analogy. Known tokens ("the water is pooling like a flood") are presented alongside partially known information ("but it is less than 2 acres and so it is not a flood"). Some features are emphasized while others are suppressed (Gernsbacher et al. 2001). This same quality has been observed in analogy (Blasko 1999, p. 1679). It informs the human that the unexpected water is pooling and rising in a discrete area, in a manner incompatible with a less urgent natural event, such as a spreading river or tide. Follow-up inferences in Stage 4 (upper right) posit that the pooling water could cause drowning.

To gain more information, the human asks for a list of items in the immediate proximity of the water. This is another use of context, but with a change to the granularity and its anchor concept (water) to better understand the immediate situation around the risk entity.

In response, the machine lists physical items that are close to the potential risk. The items listed by the machine add the important detail of the pipe and places it in contact with the water, which is also in contact with the human. These last details allow the human to create a bridging inference in which the pipe might be a source of continuous water flow. That new cluster of inferences is visualised as Stage 5. Now the human has enough information to infer the potential risk posed by the water and can make a decision to modify the plan.

The testing phases follows, but this process is concerned with the humans using explicit tokens and commands, and does not depend on implicit inference in a manner required to be illustrated as bridging inferences.

Conclusion

This research proposes a conversational framework that could be used by human-machine teams to communicate a risk condition for which there are no explicit tokens. It targets the future of human-machine collaboration, building on current limitations of machines to imagine a method based on collaborative adaptation. Current human-machine communication is restricted by the limits of the machine's ontology, as well as the difficulty of building a formal system that can adapt to open world factors such as context and novelty. An example scenario illustrates the conversational method. One goal is to strengthen adaptive capabilities in machines by making visible the level above linguistics, pragmatics and discourse, where new meaning is collectively developed. Another goal is to serve as a reference for current research into human-machine communication, so that when advanced systems are designed for the open world, the requirements for truly adaptive communication can be factored in. A final goal is to move the line between the possible and the not-yet possible, and show how the arts can contribute to human-machine interaction going forward.

Acknowledgments

This research was funded by Griffith University and Trusted Autonomous Systems, a Defence Cooperative Research Centre funded through Next Generation Technologies Fund. Thanks to Dana Kulic and Ted Goranson of Downer Defence for suggestions that could support this method.

References

Asher, Nicholas, and Alex Lascarides. 1998. "Bridging." *Journal of Semantics* 15 (1): 83–113.

Baker, Anthony L., Elizabeth K. Phillips, Daniel Ullman, and Joseph R. Keebler. 2018. "Toward an Understanding of Trust Repair in Human-Robot Interaction: Current Research and Future Directions." *ACM Transactions on Interactive Intelligent Systems* 8 (4): 1–30.

Blasko, Dawn. 1999. "Only the Tip of the Iceberg: Who Understands What about Metaphor?" *Journal of Pragmatics*, no. 31: 1675–83.

Clark, Herbert. 1977. "Bridging." In *Thinking: Readings in Cognitive Science*, 411–20. Cambridge: Cambridge University Press.

Clement, Catherine, and Karen Yanowitz. 2003. "Using an Analogy to Model Causal Mechanisms in a Complex Text." *Instructional Science* 31: 195–225.

Devlin, Keith J. 2009. "Modeling Real Reasoning." In *Formal Theories of Information: From Shannon to Semantic Information Theory and General Concepts of Information*, edited by Giovanni Sommaruga, 234–52. Lecture Notes In Computer Science 5363. Springer.

Elder, Chi-He, and Michael Haugh. 2018. "The Interactional Achievement of Speaker Meaning: Toward a Formal Account of Conversational Inference." *Intercultural Pragmatic* 15 (5): 593–625.

Elgrably, Jordan, and Milan Kundera. 1987. “Conversations with Milan Kundera.” *Salmagundi* 73 (Winter): 3–24.
Fauconnier, Giles, and Mark Turner. 2002. *The Way We Think: Conceptual Blending and the Mind’s Hidden Complexities*. New York: Basic Books.
Franke, William. 2014. *A Philosophy of the Unsayable*. Notre Dame, Indiana: University of Notre Dame Press.
French, Robert. 2002. “The Computational Modeling of Analogy-Making.” *Trends in Cognitive Sciences* 6 (5): 200–2005.
Garrett, Yanis. 2007. “Offer, Accept, Block, Yield: The Poetics of Open Scene Additive Improvisation.” Master of Philosophy, Sydney: University of Sydney.
Gernsbacher, M., B. Keysar, R. Robertson, and N. Werner. 2001. “The Role of Suppression and Enhancement in Understanding Metaphors.” *Journal of Memory and Language* 45: 433–50.
Gernsbacher, Morton Ann. 1996. “Coherence Cues Mapping during Comprehension.” In *Processing Interclausal Relationships in the Production and Comprehansion of Text*, edited by J. Costermans and M. Fayol. Hillsdale, New Jersey: Erlbaum.
Gibson, Stephen, and Cordet Smart. 2017. “Social Influence.” In *The Palgrave Handbook of Critical Social Psychology*, edited by Brendan Gough, 291–318. London: Palgrave Macmillan UK.
Goel, Ashok, Tesca Fitzgerald, and Priyam Parashar. 2020. “Analogy and Metareasoning: Cognitive Strategies for Robot Learning.” In *Human-Machine Contexts*. New York: Elsevier.
Graesser, A. C., and K. Wiemer-Hastings. 1999. “Situational Models and Concepts in Story Comprehension.” In *Narrative Comprehension, Causality, and Coherence: Essays in Honor of Tom Trabasso*, edited by S. R. Goldman, A. C. Graesser, and P. W. Van den Broek, 77–92. Mahwah, NJ: Lawrence Erlbaum.
Gruber, Thomas. 1995. “Towards Principles for the Design of Ontologies Used for Knowledge Sharing.” *International Journal of Human-Computer Studies* 43 (5–6): 907–28.
Hancock, Peter A., Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. 2011. “A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 53 (5): 517–27.
Herman, David. 2002. *Story Logic: Problems and Possibilities of Narrative*. Univ of Nebraska Press.
Irmer, Matthias. 2011. *Bridging Inferences: Constraining and Resolving Underspecification in Discourse Interpretation*. Berlin; Boston: Walter de Gruyter.
Johnson, Benjamin, Michael Floyd, Alexandra Coman, Mark Wilson, and David Aha. 2018. “Goal Reasoning and Trusted Autonomy.” In *Foundations of Trusted Autonomy*, edited by Hussein Abbass, Jason Scholz, and Darryn Reid. Vol. 117. Studies in Systems, Decision and Control. Warsaw, Poland: Springer.
Johnstone, Keith. 1981. *Impro: Improvisation and the Theatre*. Oxon: Routledge.
Kahneman, Daniel, and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica* 47 (2).
Kennedy, Daniel, and Maartje Hidalgo. 2021. “Two-Way Human-Agent Trust Relationships in Adaptive Cognitive Agent, Adaptive Tasking Scenarios: Literature Metadata Analysis.” In *Human-Computer Interaction: Theory, Methods and Tools. Proceedings 23rd HCI International Conference, HCII 2021*, edited by Masaaki Kurosu, 191–205. Lecture Notes in Computer Sc. Cham: Springer Nature Switzerland.
Kruijff, Geert, Miroslav Janicek, Shanker Keshavdas, Benoit Larochelle, and Hendrik Zender. 2015. “Experience in System Design for Human-Robot Teaming in Urban Search & Rescue.” In *Proceedings 8th International Conference on Field and Service Robots (FSR 2012)*, hal-01143147:1–15. Matsushima, Japan: Hal Open Science.
Lambert, Dale, Adam Saulwick, Chris Nowak, Martin Oxenham, and Damien O’Dea. 2008. “An Overview of Conceptual Frameworks.” DSTO-TR-2163. Defence Science and Technology Organisation.
Langacker, Ronald. 1987. *Foundations of Cognitive Grammar: Theortetical Prerequisites*. Vol. 1. Stanford: Stanford Universty Press.
Li, Jiwei, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, and Dan Jurafsky. 2016. “Deep Reinforcement Learning for Dialogue Generation.” *arXiv:1606.01541 [Cs]*, September.
O’Brien, Edward J., and Anne E. Cook. 2016. “Coherence Threshold and the Continuity of Processing: The RI-Val Model of Comprehension.” *Discourse Processes* 53 (5–6): 326–38. https://doi.org/10.1080/0163853X.2015.1123341.
Rushdie, Salman. 2008. *The Satanic Verses*. London: Random House.
Schmidt, Eric, Robert Work, Safra Catz, Eric Horvitz, Steve Chien, Andrew Jassy, Mignon Clyburn, et al. 2021. “National Security Commission on Artificial Intelligence - Final Report.” Arlington, VA USA: United States Federal Government.
Song, Yao, and Yan Luximon. 2020. “Trust in AI Agent: A Systematic Review of Facial Anthropomorphic Trustworthiness for Social Robot Design.” *Sensors* 20 (18): 5087.
Sowa, John. 2010. “The Role of Logic and Ontology in Language and Reasoning.” In *Theory and Applications of Ontology: Philosophical Perspectives*, edited by R. Poli and J. Seibt, 231–63. Berlin: Springer.
Zwaan, Rolf. 2016. “Situation Models, Mental Simulations, and Abstract Concepts in Discourse Comprehension.” *Psychon Bulletin Review* 23: 1028–34.
Zwaan, Rolf, Joseph Magliano, and Arthur C. Graesser. 1995. “Dimensions of Situation Model Construction in Narrative Comprehension.” *Journal of Experimental Psychology* 21 (2): 386–97.