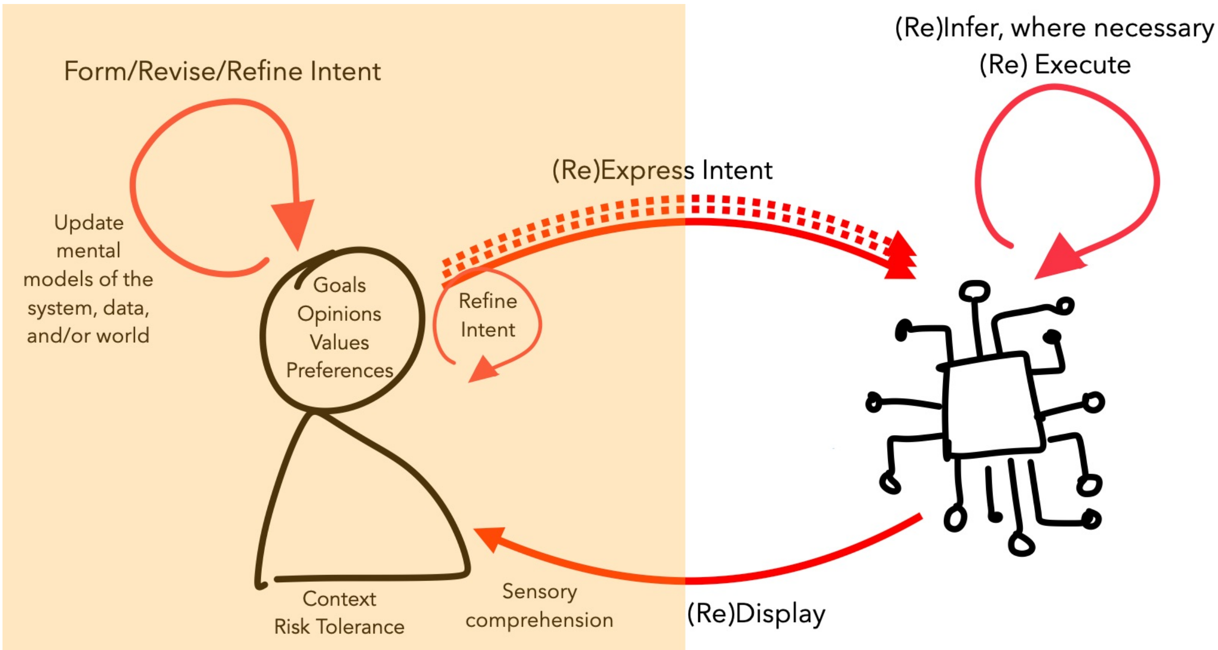

Figure 1: A snapshot of the framework in its current form at a high level

A Framework for Designing Human-Computer Communication Interfaces

Elena L. Glassman, Harvard University SEAS

Keywords

frameworks, human-computer communication, interface design, transformative reflection

Intent

In order to communicate something to someone else—or some*thing* else—it is necessary to have something in mind to be communicated, i.e., an intent. It could be

- a category (e.g., binary classification),
- a function (e.g., a personal similarity metric),
- a target image in the mind,
- a notion of something to say in natural language (e.g., an email that says no politely),
- an operation (e.g., FlashFill [4]), or
- a program.

This intent may originate initially from the person, as a natural consequence of their goals, opinions, values, preferences, and context, or it may be a reaction to a system's display, e.g., a representation or example of its capabilities, a dataset, or a view of the world.

Expression

The human can express their intent using one or more modalities at the same time, by

- composing statements in a language—a natural and/or programming language—including program sketches;
- providing concrete examples, demonstrations, or partially-concrete linguistic sketches;
- annotations of whatever is already in or has just been added to the *common ground* shared by the system and the human (e.g., RAGAE [8]);
- physical gestures, e.g., in open space or on a touch sensitive surface; and
- GUI interactions, e.g., physical and virtual button pushes that invoke a particular function.

Even for an unchanged intent, the human might even try multiple separate ways to specify what they want in parallel, in case one way of expressing what they want is more effective for the system than another, if the system can work on each specification in parallel threads; this can minimize the human's anxiety about making mistakes in any one specification they provide in more complex domains, e.g., synthesizing programs in a language they do not know [5].

In the process of expressing their intent, the human may have new insights about their intent that impact their intent and their intent expression even before receiving a response from the system.

Relevant concepts in the historical literature include the Gulf of Execution, referring to when the user struggles to use the affordances given to them to express their intent such that the system correctly interprets them.

Inference and Execution

If the user expresses their intent in a way that requires no inference, e.g., as statement(s) in a programming language or as a push of a button that invokes a pre-programmed function, then the system can just execute the expressed intent and reflect any feedback to the user in the next step.

If the intent expression has any semantic ambiguity, it is necessary for the system to incorporate some AI/ML to perform inference about the intended intent. There are multiple types of potential inference errors, such as mistaking one spoken word for another or misinterpreting the semantic meaning of a correctly transcribed natural language utterance.

Feedback

Feedback to the user can include:

- user-relevant components of the system's state, e.g., how many alarms have been set (see also the ”Visibility of System State” recommendation within Nielson's usability heuristics);
- any inferences the system made based on the user's intent expression, any model it has of the user, and its own base priors and heuristics;
- a view (or preview) of what executing the inferred request does, given real or hypothetical data or situations

This feedback becomes part of the common ground shared with the human user and, depending on how it is provided to the user, can be explicitly annotated or edited as part of subsequent rounds of intent expression.

Attention, Comprehension, and Mental Modeling

While this feedback can be delivered, it may or may not be received by the human due to issues with attention and sensory-level comprehension:

- not noticing the information, e.g., visual information being too far away from where they are looking
- not noticing the information due to it being encoded in a way they cannot perceive, e.g., due to color-blindness
- only delivering information on one channel that is blocked, e.g., audio feedback when a device has been muted

If the feedback is received, it still needs to be interpreted and comprehended. For example, the human needs to actively construct the meaning of any feedback delivered in visual, natural, or programming languages.

As a result of this comprehension, the human may consciously or unconsciously update their mental models of:

- the system,
- any data at hand, and
- the world

in which the system might act in/on now or in the future. Based any mental model updates and personal factors listed previously, as well as task-specific risk tolerances, the human may refine or entirely revise the intent that drives any subsequent intent expressions.

Evaluation

Within this conversational loop, the human has the greatest access to their own goals, values, preferences, and context; as a result, only they can decide when the system has sufficiently understood and can correctly carry out the final version of the intent they have attempted to communicate.

In additional to traditional measures of usabilty evaluation, system designers can literally count the number of trips around this conversational loop it takes for the human to reach the point of the human confidently and correctly understanding that the system has understood their intent.

In situations where the system is helping the human make a decision or construct an object, the human may more quickly or better fulfill their goals by authoring the final result themselves in the process of or as a result of interacting with the system, even though the system never correctly understood their intent. Interfaces can explicitly afford this, and when counting conversational loops, this is an alternative place to end. User studies that force the user to keep going, expressing and re-expressing their intent, are misleading.

Additional Cognition Considerations

Consuming more information also requires time and cognitive resources that we know are scarce. Our cognition is effortful and limited. Like our natural avoidance of pointless physical exertion [6], we conserve our mental energy. Our ability to hold and manipulate complex situations and ideas can be expanded in any particular domain through specialized training or augmented with specific tools but will always have its limits. To help us make a decision despite this, we unconsciously deploy heuristics and biases to make *some* judgment, despite even substantial remaining uncertainty [7], so we can move on with our lives rather than being frozen in indecision. Recent work on cognitive engagement & incidental learning [3], as well as the impacts of cognitive forcing functions [1, 2], speak to these concerns within human-AI interaction specifically.

References

[1] Z. Buçinca, P. Lin, K. Z. Gajos, and E. L. Glassman. Proxy tasks and subjective measures can be misleading in evaluating explainable ai systems. In Proceedings of the 25th international conference on intelligent user interfaces, pages 454–464, 2020.

[2] Z. Buçinca, M. B. Malaya, and K. Z. Gajos. To trust or to think: cognitive forcing functions can reduce overreliance on ai in ai-assisted decision-making. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1):1–21, 2021.

[3] K. Z. Gajos and L. Mamykina. Do people engage cognitively with ai? impact of ai assistance on incidental learning. In 27th International Conference on Intelligent User Interfaces, pages 794–806, 2022.

[4] S. Gulwani. Automating string processing in spreadsheets using input-output examples. ACM Sigplan Notices, 46(1):317–330, 2011.

[5] J. Hu, P. Vaithilingam, S. Chong, M. Seltzer, and E. L. Glassman. Assuage: Assembly synthesis using a guided exploration. In The 34th Annual ACM Symposium on User Interface Software and Technology, pages 134–148, 2021.

[6] D. Lieberman. Exercised: Why something we never evolved to do is healthy and rewarding. Vintage, 2021.

[7] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. Science, 185(4157):1124–1131, 1974.

[8] T. Zhang, L. Lowmanstone, X. Wang, and E. L. Glassman. Interactive program synthesis by augmented examples. In Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, pages 627–648, 2020.