

Validating Interaction Patterns in HRI

Peter H. Kahn, Jr.¹, Brian T. Gill², Aimee L. Reichert¹, Takayuki Kanda³, Hiroshi Ishiguro^{3,4}, & Jolina H. Ruckert¹

¹Department of Psychology
University of Washington
Seattle, Washington, USA
pkahn@uw.edu
aimeer3@uw.edu
jhr333@uw.edu

²Department of Mathematics
Seattle Pacific University
Seattle, Washington, USA
bgill@spu.edu

³Intelligent Robotics and
Communication Laboratory
ATR
Kyoto, Japan
kanda@atr.jp

⁴Department of Systems
Innovation
Osaka University
Osaka, Japan
Ishiguro@sys.es.osaka-u.ac.jp

Abstract— In recent work, “interaction patterns” have been proposed as a means to characterize essential features of human-robot interaction. A problem arises, however, in knowing whether the interaction patterns generated are valid. The same problem arises when researchers in HRI propose other broad conceptualizations that seek to structure social interaction. In this paper, we address this general problem by distinguishing three ways of establishing the validity of interaction patterns. The first form of validity seeks to establish whether the conclusions about interaction patterns are warranted from the data. The second seeks to establish whether the interaction patterns account for the data. And the third seeks to provide sound reasons for the labels of the patterns themselves. Often these three forms of validity are confused in discussions about conceptual categories in HRI.

Keywords-human-robot interaction, interaction patterns, validity

I. INTERACTION PATTERNS IN HRI

Interaction patterns is an emerging approach for assessing sociality in HRI [1]. An interaction pattern characterizes essential features of social interaction between humans and robots, specified abstractly enough such that many different instantiations of the pattern can be uniquely realized given different types of robots, purposes, and contexts of use. Consider the following example of an interaction pattern between humans: When we are introduced to a new person, we often shake hands, say ‘hi,’ exchange our names, and perhaps engage in a little chit-chat about the weather. In other cultures, people might bow to one another, or offer a Namaste greeting. We can call this universal social activity an “Introduction.” The Introduction facilitates further social interaction. While the Introduction is never enacted exactly the same way twice, the activity is structured, it follows a recognizable pattern. And this pattern is but one of many that can be used to help structure human-robot interaction.

In 2008 we presented an initial account of how the Interaction Pattern approach can be implemented in a laboratory context [1]. At that time, we proposed 8 possible human-robot design patterns, though have now recast the approach as “interaction patterns” to highlight the interactional aspect of what we seek to capture. These patterns were: (i) introduction, (ii) didactic communication, (iii) in motion together, (iv) personal interests and history, (v) recovering from mistakes, (vi) reciprocal turn-taking in game context, (vii)

physical intimacy, and (viii) claiming unfair treatment or wrongful harms.

While we believe our approach toward characterizing social interaction in HRI based on the structure of interaction patterns is promising, and while we are moving the approach forward in our current research, we have encountered the following question: “Are the proposed interaction patterns valid?” This paper offers our first iteration toward answering that question. To do so, we distinguish three ways of establishing the validity of interaction patterns.

II. THREE WAYS TO CONCEIVE OF VALIDITY

Perhaps the most common way researchers conceive of validity is in terms of whether conclusions are warranted from the data [2]. Psychometrics is perhaps the most common field that is then drawn upon to establish validity of this form. For example, in psychological and educational testing, one might collect data on school performance and IQ, and attempt to validate the conclusion that IQ predicts school performance. It is in this sense that IQ would be said to be or not to be (depending on the results) a valid measure. Notice that it is not the case that IQ itself is valid or not valid. That is not a sensible proposition [3]. Rather, the proposition is that IQ can be used to accurately predict something else, and once that something is specified then whether that conclusion is warranted can be assessed.

Often such assessments entail quantitative analyses. But in principle, this form of validity can be established qualitatively, as well. For example, in the field of Literature, one might collect “data” from the narrative text of Shakespeare’s *King Lear* and argue that the play is substantially about the difficulty old men have in giving up power and of recognizing love in family settings. It can then be asked, is that interpretation valid? That is a sensible question, and it requires the literary theorist to provide textual evidence and good reasons for the interpretation. It is in this way – in terms of establishing whether a conclusion is warranted based on the data – that asking “Is one’s interpretation of *King Lear* valid?” is of the same form of validity as asking “Is IQ a valid predictor of school performance?”

Interaction patterns can be subject to this form of validation. But it does not suffice to ask such questions as: “Does the interaction patterns work?” or “Are the interaction patterns useful.” For such words as “work” and “useful” need to be further specified in terms of the conclusions being put

forward. Such a conclusion might be that the patterns are effective in facilitating human-robot interaction, which is amenable to this first form of validation.

A second form of validating an interaction pattern can be framed in terms of the following question: “Does the pattern account for the data?” This form of validation is less common to how someone in psychometrics views validity, and it is more common to how a modeler views validity. More specifically, we propose that interaction patterns need to “fit the data” in terms of form and function. As a case in point, the form of the introduction pattern in a dyadic context involves a constellation of behavioral interactions: for example, people physically meeting, engaging in short exchanges of salutation (“hello”), which are responded to by the other (“hi”), ritualized statements of initial pleasure (“pleased to meet you”), ritualized questions of concern for the other (“how are you?”), physical reciprocity, often with hands (handshake or namaste greeting) or with the body (bowing), with bodies frontally positioned to one another, and so forth. Interaction patterns that account for more data are, in this sense of the term, more valid than interaction patterns that account for less data.

This form of validation helps to address a concern that some people have voiced informally about our research agenda on Interaction Patterns. The critique is that while the interaction patterns we propose can make sense, the patterns seem very subjective and that we appear to preclude hundreds of other patterns that could be generated. But the subjectivity of our patterns are checked by this form of validation insofar as the patterns need to account for actual data. If the patterns do, they are valid in this sense of the term. It may well be, then, that hundreds of other patterns exist. But that answer depends on the data yet to be accounted for.

The third form of validation seeks to provide sound reasons for the labels of the patterns themselves. Why, for example, call the Introduction an Introduction? Why not call it a “Greeting”? or a “Meeting”? or a “Gathering” or a “Lunch-Time Break”? To develop our answer, think of the distinction between formal and informal fallacies in logic. Formal logical fallacies occur through errors in deductive reasoning (e.g., if $a < b$, and $b < c$, then it is a fallacy not to agree that it must follow that $a < c$). In contrast, informal logical fallacies occur when bad qualitative reasons are offered (e.g., in “argumentum ad hominem” it is a fallacy to critique someone’s argument based on the person’s character). It is in this informal sense that interaction patterns, to be valid, need good reasons for being called what they are called. Some of those reasons will follow standard understandings of what words mean in one’s established language (e.g., English). Other reasons can be established based on philosophically grounded arguments for conceptualizing a construct. As a case in point, some moral developmental psychologists have drawn on moral philosophers such as John Rawls to characterize the moral domain in terms of judgments that pertain to human welfare and fairness, and which generalize to people cross-culturally [4]. In other words, well-grounded arguments in philosophy have helped to establish the validity of constructs in psychology.

As the HRI field evolves, many useful approaches for characterizing social interaction between humans and robots will involve conceptual categories. For example, in recent years researchers in HRI have proposed different forms of conceptual “benchmarks” to guide the design of social robots. Benchmarks have been defined as, “categories of interaction that capture fundamental aspects of human life”, and to date include autonomy, imitation, intrinsic moral value, moral accountability, privacy, reciprocity, conventionality, creativity, and authenticity of relation, intersubjectivity, safety, scalability, autonomy, imitation, privacy, understanding of domain, social success, and impact on the user’s care, life and caregivers , and relationship potential [5].

III. CONCLUSION

The *Oxford English Dictionary* says the following about validity, that it is the “quality of being well-founded on fact, or established on sound principles, and thoroughly applicable to the case or circumstances; soundness and strength (of argument, proof, authority, etc.).” In turn, there are different ways to establish this quality of being well-founded. In this paper, we have offered three such overarching ways in the context of interaction patterns. To the extent that interaction patterns are one example of conceptual categories in HRI, our approach toward establishing validity may well turn out to have broad reach.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. IIS-0842832. 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 National Science Foundation. We thank Greg Trafton and Daniel Levin for early discussions on these matters.

REFERENCES

- [1] P. H. Kahn, Jr., N. G. Freier, T. Kanda, H. Ishiguro, J. H. Ruckert, R. L. Severson, and S. K. Kane, “Design patterns for sociality in human-robot interaction,” in Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction (Amsterdam, The Netherlands, 2008), New York, NY: ACM Press, 2008, 97-104.
- [2] L. J. Cronbach, “Test validation,” in Educational Measurement, 2nd ed., R. L. Thorndike, Ed., Washington, DC: American Council on Education, 1971, 443-507.
- [3] S. Messick, “Validity of psychological assessment,” *American Psychologist*, 50, 1995, 741-749.
- [4] E. Turiel, “The Development of Morality,” in *Handbook of Child Psychology. Social, Emotional, and Personality Development*, 5th ed., vol. 3, W. Damon and N. Eisenberg, Eds., New York, NY: J. Wiley, 1998, 863-932.
- [5] P. H. Kahn, Jr and K. F. MacDorman, “Special issue: Psychological benchmarks of human-robot interaction,” *Interaction Studies*, 8:3, 2007, 359–517.