Proceedings Second International Symposium on New Frontiers in Human-Robot Interaction, K. Dautenhahn, J. Saunders (Eds.), at the AISB 2010 convention, 29 March – 1 April 2010, De Montfort University, Leicester, UK

Validating Characterizations of Sociality in HRI: The Case of Interaction Patterns

Peter H. Kahn, Jr. [1], Brian T Gill[2], Aimee L. Reichert[1], Takayuki Kanda[3], Hiroshi Ishiguro[3,4], and Jolina H. Ruckert[1]

Abstract. As the HRI field evolves, researchers increasingly seek to provide characterizations of sociality in human-robot interaction. But how does one assess whether the characterizations are valid? Using our research on "interaction patterns" as a case in point, this paper offers 5 approaches toward establishing validity: The psychometric approach, literary approach, modeling approach, philosophical approach, and structural approach. We argue that when it comes to validating characterizations of sociality in HRI, too often people ask for evidence of psychometric validity, without clarity of what that involves, and without awareness of the benefits of these other approaches.

1 INTRODUCTION

As the HRI field evolves, researchers increasingly seek to provide characterizations of sociality in human-robot interaction. These characterizations may be framed broadly, as when researchers seek to show how people establish *friendships* with robots, or *trust* robots, and *care* about robots, or attribute to robots *moral accountability* [6, 13, 17, 24, 34, 37, 46, 48]. The characterizations may also be framed with greater specificity. For example, researchers in HRI have proposed different characterizations of psychological "benchmarks" to measure success in building increasingly humanlike robots. To date, researchers have characterized such benchmarks as *privacy*, *reciprocity*, *imitation*, *intrinsic moral value*, *conventionality*, *creativity*, *authenticity of relation*, and *autonomy* [5, 9, 15, 33, 37, 51]. Characterizations such as these are "foundational" in the sense that they seek to say something substantive, deep, and abiding about human-robot interaction.

When researchers propose foundational characterizations of sociality in HRI, an important question emerges: How does one know whether the characterizations are valid? It is an important question for if one wants to know something about the world, one almost always wants one's knowledge to be valid. The problem however with this question is that it is often asked as if the answer will and indeed must entail some psychometric assessment of validity. But the psychometric approach to validity is only one of many, and not often the most amenable to validating characterizations of sociality.

[1]Dept. of Psychology, Univ. of Washington, Seattle, Email: {pkahn, aimeer3, jhr333}@uw.edu.
[2]Dept. of Mathematics, Seattle Pacific Univ., Email: bgill@spu.edu.
[3]Intelligent Robotics and Communication Lab, ATR, Kyoto, Japan, Email: kanda@atr.jp.
[4]Dept. of Systems Innovation, Osaka Univ., Japan, Email: Ishiguro@sys.es.osaka-u.ac.jp

Thus in this paper we distinguish between five meanings of validity, and their corresponding methods, using our research on "interaction patterns" as a case in point. Our purpose is to provide some clarity when researchers seek to assess the validity of their proposed characterizations of sociality in HRI.

2 INTERACTION PATTERNS

By interaction patterns we mean foundational features of social interaction between humans and robots, characterized abstractly enough such that many different instantiations of the pattern can be uniquely realized given different types of robots, purposes, and contexts of use, while delimited so as to meaningfully segment interaction [31].

Figure 1. Initial Introduction Interaction Pattern

Consider, for example, the interaction pattern we have called the "Initial Introduction" (Figure 1). When people are introduced to a new person, typically in a Western culture they engage in a largely scripted and conventionally-established interactional repertoire wherein they (a) recognize the other with their eyes, (b) engage in ritualized short exchanges of salutation ("hello"), (c) which are responded to by the other ("hi"), (d) offer

reciprocal statements of initial pleasure ("pleased to meet you", "pleased to meet you, too"), (e) ask polite questions of concern for the other ("how are you?"), and (f) engage in physical reciprocity (a handshake), (g) with bodies frontally positioned to one another. In non-Western cultures, different instantiations of this pattern occur. For example, people might bow to one another upon meeting, instead of shaking hands, or offer each other a Namaste greeting with hands together. While the Introduction is never enacted exactly the same way twice, the activity is structured, it follows a recognizable pattern. It is likely universal. If we were dropped into an unknown culture, it is likely that people in the culture would enact some form of this pattern, and that when we saw it enacted for the first time we would immediately recognize it. In turn, such a pattern is but one of many that we believe can be used to help structure human-robot interaction.

Our conceptual framework for interaction patterns draws on Christopher Alexander and his colleagues work in the field of architecture. According to Alexander et al. [4], a design "pattern describes a problem which occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over, without ever doing it the same way twice" (p. x). For example, one of their 253 patterns is titled "Light on two sides of every room." They write: "The importance of this pattern lies partly in the social atmosphere it creates in the room" (p. 748). They also write that this "pattern, perhaps more than any other single pattern, determines the success or failure of a room" because "when they have a choice, people will always gravitate to those rooms which have light on two sides, and leave the rooms which are lit only from one side unused and empty" (p. 747). There is now a burgeoning body of work that extends the idea of design patterns into the fields of ubiquitous computing [10], software engineering [18], interaction design [7, 45], and usability [8, 25].

In 2008 we presented an initial account of how the Interaction Pattern approach can be implemented in a laboratory context, and provided descriptions of eight such patterns [31]. The Introduction is one of those. We now summarize the other seven interaction patterns to provide the reader with a better sense of what they look like.

Didactic Communication. One of the least complicated forms of social communication involves the transmission of information from one to another. Didactic Communication is an interaction pattern for such one-way communication of information, situated in a context where each party has motivation to remain engaged. When teachers lecture to a class they embody this interaction. Potentially robots, even with today's capabilities, could be very good at generating instantiations of this interaction since it requires minimal responsiveness to the listener.

Prosocial Request. Part of social interaction involves at times requesting assistance from another: for example, asking for the whereabouts of an item, for directions to a location, or for how to use a tool or new software program.

In Motion Together. Being in a social relationship with others can involve aligning one's physical movement with others, such as often occurs when walking together (Figure 2).

Personal Interests and History. Social life gains much of its character and depth through sharing of one's personal interests and history with others, and of being known and knowing others through such sharing.

Recovering from Mistakes. Social life involves at times the making of mistakes. Recovering from Mistakes is an interaction pattern that creates the potential for both parties to maintain a social affiliation following the mistake.

Physical Intimacy. One important social characteristic of humans (and most primates) is that they engage in holding, touching, embracing, and other forms of physical intimacy. Physical Intimacy is an interaction pattern that allows this characteristic to find expression.

Turn-Taking in Game Context. Most social games, no matter their variation, involve taking turns with one another, such as many board games (e.g., Monopoly), many card games (e.g., poker), tag, and baseball, to name just a few. Turn-Taking in a Game Context is an interaction pattern for sociality that may easily set into motion claims of unfairness.

Figure 2. In Motion Together Interaction Pattern

If an interaction patterns program – something of the form we are proposing here – proves successful, it could provide HRI researchers with (a) a new methodology by which to conceptualize and build meaningful social relationships with robots, (b) foundational basic knowledge about what is unique to human-robot interaction, as compared to human-human interaction, and (c) a more systematic approach to help shape decisions about which technical problems in HRI to work on insofar as certain interaction patterns will likely be found to be particularly important in supporting meaningful social relationships, and thus might well merit greater resources (funding and time) to find the means for robots to implement the interaction patterns autonomously in different contexts.

3 FIVE APPROACHES TO VALIDATING CHARACTERIZATIONS OF SOCIALITY

With this handful of interaction patterns as a backdrop, we can now turn to different ways that validity can be understood and assessed.

3.1 The Psychometric Approach

Perhaps the most common approach that people think of in terms of establishing validity is based on psychometrics. But what this approach is about has changed as ideas about psychometric validity have taken shape for over sixty years.

In 1954, the American Psychological Association (APA) published its first *Technical Recommendations for Psychological Tests and Diagnostic Techniques* [1]. This publication described four categories of validity: *predictive validity*, *concurrent validity*, *content validity*, and *construct validity*. In 1966, the APA with the American Educational Research Association and the National Council on Research in Education together published an update [2]. The 1966 update combined predictive and concurrent validity into a single category called *criterion-related validity*. Criterion-related validity involved comparing scores on a test or scale with other variables or measures taken on the same subjects, typically by computing correlations between the different measures. If the different measures were all obtained at approximately the same time, criterion-related validity was known as concurrent validity. Concurrent validity was often studied to determine whether or not one test could substitute for another; or, in other words, whether or not different tests were measuring "the same thing." When the score on the test was compared to other measures obtained some time after the original test, criterion-related validity became known as predictive validity.

In turn, in this same 1966 landmark publication, content validity was understood as assessing how well the content of a test actually represented the entire range of situations or subject matter about which conclusions were to be drawn. Assessments of content validity were somewhat subjective and not quantified with statistical methods; but some rational justification of content validity was necessary to avoid easy rejection of a researcher's findings. For example, a researcher might develop a survey consisting of questions relating to issues such as abortion and gay rights. On the basis of this survey, the researcher might develop a score which was claimed to represent how liberal or conservative respondents were. However, this interpretation of the scores would presumably have poor validity since the questions used on the survey only represent highly contentious social issues and not a wider range of subject matter (such as economic issues) presumably implied by the labels "liberal" and "conservative."

In 1966, construct validity was also discussed extensively, though the term did not mean quite what it sounds like. According to Cronbach [12] "construct validity is evaluated by investigating what psychological qualities a test measures; i.e., by determining the degree to which certain explanatory concepts or constructs account for performance on the test" (p. 144). As an example, consider the IQ test which seeks to provide a quantitative measure of the construct of intelligence. Is intelligence a valid construct? Construct validation does not actually answer that question. Rather, the focus is on the interpretation of the scores on the IQ test itself, of whether the interpretations are warranted based on empirical evidence.

The emphasis on the different "types" of validity in the 1950's and 1960's led to a fragmented view of psychometric validity, and there was eventually a movement to try to unify the theory of validity. Ultimately, the goal of all psychometric validation questions is to assess "the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests" (p.9) [3]. Over time, all test scores came to be viewed as measures of some underlying construct, and the first four forms of validity (content, predictive, concurrent, and criterion-related validity) came to be viewed as aspects of construct validity. Since it is the link between the test scores and the underlying construct that gives the scores their meaning and supports their interpretation, ultimately all psychometric validity questions are construct validity questions. Thus in their most recent Standards [3], the American Psychological Association dropped the term construct validity and reverted back to simply the term validity. For example, the most recent American Psychological Association position [3] states:

> Validation logically begins with an explicit statement of the proposed interpretation of test scores, along with a rationale for the relevance of the interpretation to the proposed use. The proposed interpretation refers to the construct or concepts the test is intended to measure. Examples of constructs are mathematics achievement, performance as a computer technician, depression, and self-esteem. To support test development, the proposed interpretation is elaborated by describing its scope and extent and by delineating the aspects of the construct that are to be represented. The detailed description provides a conceptual framework for the test, delineating the knowledge, skills, abilities, processes, or characteristics to be assessed. The framework indicates how this representation of the framework is to be distinguished from other constructs and how it should relate to other variables (p. 9).

With such language, all of the different aspects of validity are being wrapped within the single term. Rather than talking about different "types" of validity, the APA intentionally shifted toward emphasizing different kinds of evidence that might be called upon to support interpretations of test scores for particular uses. Thus, rather than using terms such as "content validity" or "criterion-related validity", the 1999 Standards discuss "evidence based on test content" and "evidence based on relations to other variables." All of the available evidence is then combined together to form a single, coherent argument for the validity of the desired interpretations and uses of the test scores.

One key idea stated in the above APA definition is that it is not a sensible question to ask whether a test is valid or not valid. Rather, it is the interpretation of the test scores that are valid or not valid. As the passage states above: "Validation logically begins with an explicit statement of the proposed interpretation of test scores, along with a rationale for the relevance of the interpretation to the proposed use." Psychometricians have emphasized this idea time and time again since the 1950's, because people tend to think otherwise. They tend to think that tests themselves are valid or not.

To emphasize this idea – that it is interpretation of tests, and not the test itself, which is the subject of validation – let's imagine that we develop a new standardized test of basic algebra skills. Now let's say that three different schools use our test in three different ways. One school uses our test to assess the effectiveness of instruction in their algebra courses. The second school uses our test as a screening tool to determine whether or not to allow students into a geometry course, which is next in the math sequence. The third school uses our test to plan instruction in a course by giving the test to students at the beginning of the course and using an item-by-item analysis of student errors on the test to select topics to focus on during the course. Each of these different uses of the test involves different interpretations and conclusions based on students' performance on the test. In turn, our test might prove to be an excellent assessment tool for the end of the algebra course. But our test may correlate very poorly with success in the sequenced geometry course, and it would likely enough function very poorly as the basis for an item-by-item restructuring of curriculum. If that was what we found empirically, then it would be valid to use the test in the first way, but not in the second and third.

These strengths and constraints of psychometric validity apply not just to scores on a test but "generically in its broadest sense to mean any coding or summarization of observed consistencies or performance regularities on a test, questionnaire, observation procedure, or other assessment devices such as work samples, portfolios, and realistic problem simulations" (p. 741) [40]. It follows, then, that questions such as "Is the questionnaire valid" or "is your observation procedure valid?" or "is your realistic problem simulation valid?" are not sensible questions from a psychometric standpoint [39]. No more so than "Is the test valid?"

And that is precisely the point we want to drive home about validity and interaction patterns. From the psychometric approach to validity, it is not sensible to ask whether an interaction pattern is valid or not valid. No more so than "Is the test valid?" Or "Is the questionnaire valid?"

The one caveat is that it is possible to invoke the conception of content validity, described above, and ask in that sense of the term whether an interaction pattern is valid. But as we discussed above, evidence for content validity is not based on quantitative or statistical analysis, and tends to be viewed by the psychometric field as not being rigorous. To make it rigorous, we think some of the other approaches to validity described later in this paper could be better employed.

Before we conclude this section on psychometric validity, a reader might respond: "Well, what I mean by the term psychometric validity, as it applies to an interaction pattern is: Does the interaction pattern work? Is it useful? If it does and is, then it's valid; otherwise it's not." These are reasonable questions. But they are not yet framed in a way that is tractable from the psychometric approach. To reframe them, we would need to specify what is meant by "work" or "useful" in terms of interpretations or conclusions that might follow from the interaction pattern. For example, let's say that we first hypothesized that the Introduction pattern facilitates the ease of an initial greeting. Let's say we could find ways of measuring the "ease of an initial greeting." Then let's say we compared the Introduction pattern to other patterns or entities to test which provides the greatest ease of an initial greeting. At this point, the interaction pattern itself becomes a factor in an experimental design. And, in this sense, depending on the empirical results, the interaction pattern could be said to be valid or not. But notice now that we are right back to the essence of what is meant by psychometric validity: whether interpretations of a test score or comparable entity (in this case interaction patterns) are warranted based on the empirical data. That is fine. But we still have not answered the question of whether the characterization of the phenomenon in terms of the specification of the pattern is in itself valid. For that, we need to draw on other validation approaches.

3.2 The Literary Approach

It might seem that that the fields of psychometrics and literary theory have little in common, especially in terms of what they have to say about validity. But both approaches seek to answer a fundamental question: "Are the conclusions warranted from the data?"

In literature one might, for example, 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; or argue more nuanced positions. It can then be asked, are those interpretations valid? That is a sensible question, and it requires the literary theorist to provide textual evidence and good reasons for the interpretations. 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?"

This point is important to recognize. As mentioned earlier, psychometricians often discount the rigor involved in establishing content validity, but that is likely because they focus on questions that have quantitative answers. But literary theorists know all too well that their positions require detailed, sophisticated, logical, and well-documented arguments, and counterarguments, to hold sway and to be valid. For example, literary theorist Stanley Fish highlights the distinction between interpretation and textual evidence, he writes: "The rhetoric of critical argument, as it is usually conducted in our journals, depends upon a distinction between interpretations on the one hand and the textual and contextual facts that will either support or disconfirm them on the other…" (p. 340) [16]. Herbert Dingle [14], on the scientific approach and its application in literary criticism quotes R. G. Moulton: "The fundamental axiom of inductive literary criticism is that '*Interpretation in literature is of the nature of a scientific hypothesis, the truth of which is tested by the degree of completeness with which it explains the details of the literary work as they actually stand*'" (pp. 6-7 italics part of original text). E.D. Hirsch [28] expands upon this approach and presents a model for establishing the validity of an interpretation. He writes:

> To establish a reading as probable it is first necessary to show . . . that it is possible. This is the criterion of legitimacy: the reading must be permissible within the public norms of the langue [Saussure's term for the language as a system] in which the text was composed. The second criterion is that of correspondence: the reading must account for each linguistic component in the text. Whenever a reading arbitrarily ignores linguistic components or inadequately accounts for them, the

> reading may be presumed improbable. The third criterion is that of generic appropriateness: if the text follows the conventions of the scientific essay, for example, it is inappropriate to construe the kind of allusive meaning found in casual conversation. When these three preliminary criteria have been satisfied, there remains a fourth criterion which gives significance to all the rest, the criterion of plausibility or coherence. The three preliminary norms usually permit several readings, and this is by definition the case when a text is problematical. Faced with alternatives, the interpreter chooses the reading which best meets the criterion of coherence (p. 1406).

Here Hirsch presents the methods by which to establish an interpretations' validity, which includes careful observations and logical inferences. An interpretation that captures the greatest coherence or plausibility among the elements within the text is judged to be the most valid [21]. Over the last four decades, this approach has been extended into the social sciences, often under the rubric of "hermeneutics," wherein human behavior has been treated as a "text" to be interpreted, and validity established along similar lines [19, 29, 38].

Thus the literary approach to validity is just as demanding and rigorous as the psychometric approach, though qualitative and not quantitative in form. Moreover, this approach emerged as its own field, as hermeneutics, in terms of applying it to the social sciences. Accordingly, it is possible that some degree of validation of interaction patterns can occur by drawing on this literary approach, where the "text" is the human-robot interaction being interpreted, and where the interaction pattern then becomes in effect the "conclusion" of an argument.

3.3 The Modeling Approach

A third way of understanding the validation of interaction patterns can be framed in terms of the following question: "How well does the pattern bear up compared to the corresponding human benchmark?" Benchmarks in HRI have been defined as "categories of interaction that capture fundamental aspects of human life" [37]. This form of validation is less common to a psychometric approach to validity, and more akin to the approach of a social science modeler. Such a modeler attempts to represent theory about some complex real-world phenomenon by simulating components of the phenomenon [47]. The plausibility of the model increases – and thus it becomes more valid – when model performance matches real-world findings [49].

As a case in point, consider the embodied learning model of infant gaze-following developed by Trafton et al. [49]. Their model simulated 5 theoretical components thought to be central to the emergence of gaze-following in infants. These components were reactivity, habituation, spatial module, gaze-following, and utility learning. To test the plausibility of their model they matched their model's performance to results from an experiment by Corkum and Moore [11], which assessed gaze-following in 6-11 month old infants. The embodied gaze-following model matched these results, and thus provided evidence for the validity of their model. In this sense, the empirical data served as a human-benchmark for gaze-following.

Similarly, interaction patterns can be used to better understand the complexities of social interaction, and they can be said to be more plausible – and thus more valid – when the patterns incorporate components that maximize the match to human-human interactions. That said, an important distinction between modeling and interaction patterns is that a model's validity increases when a parsimonious understanding of real-world systems is achieved by focusing on the least number of components that can account for the real-world data [49]. For example while Trafton et al.'s [49] gaze-following model initially incorporated 5 components, they found that only three of them were critical to the success of the model. In contrast, one way of understanding interaction patterns is that they seek to maximize the number of components accounted for by the pattern in order to better represent the diversity of social interaction.

Recall, for example, the specific details of the Introduction pattern described earlier. We posited that when people enact this pattern, they typically recognize the other with their eyes, engage in ritualized short exchanges of salutation that are responded to by the other, offer reciprocal statements of initial pleasure, ask polite questions of concern for the other, and engage in physical reciprocity, with bodies frontally positioned to one another. It may be possible to enact this pattern with only a single component, such as the ritualized short exchange ("Hello"), and still fit the pattern to human-human interaction, but the simplification of the pattern in this way fails to capture the rich and diverse social interplay that is possible – and often enacted – between interacting entities. Thus interaction patterns that account for more components, while still falling within the parameters of the pattern itself, are in this sense of the term more valid than interaction patterns that account for fewer components.

3.4. The Philosophical Approach

The fourth form of validity 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 [26]. As a case in point, some moral developmental psychologists have drawn on moral philosophers such as John Rawls [44] to characterize the moral domain in terms of judgments that pertain to human welfare and fairness [30], and which generalize to people cross-culturally [35, 50]. In other words, well-grounded

arguments in philosophy have helped to establish the validity of constructs in psychology.

Because we have now invoked the word "construct," the reader might suggest that construct validity is best established through a psychometric approach rather than a philosophical approach. But as we broached in Section 3.1, it is important to understand what construct validity is and what it is not from a psychometric standpoint. The 1966 APA publication, which, as noted earlier, is one of the canonical standards on validity, says the following:

> [Construct validity] is evaluated by investigating what qualities a test measures, that is by determining the degree to which certain explanatory concepts or constructs account for performance on the test. To examine construct validity requires a combination of logical and empirical attack. Essentially, studies of construct validity check on the theory underlying the test. … Construct validity is ordinarily studied when the tester wishes to increase his understanding of the psychological qualities being measured by the test. …. Construct validity is relevant when the tester accepts no existing measure as a definitive criterion of the quality with which he is concerned…or when a test will be used in so many diverse decisions that no single criterion applies (p.13).

We find this key passage, and others like it, difficult to interpret. On the one hand, construct validity, by the sound of its name, should seemly tell us whether it is valid or not valid to conceptualize constructs such as intelligence as a unified whole. On the other hand, the above passage is mostly focused on the test (e.g., whether "constructs account for performance on the test" and "the psychological qualities being measured by the test"). It does not actually say anything specific about the theoretical construct itself. The passage does not show us the means, for example, of whether we should hold out for a single construct called intelligence, or whether, as others have proposed [20] there are multiple intelligences, such as mathematical, musical, linguistic, artistic, and social, or whether all of these constructs are not framed in a valid way. It is for this reason that the philosophical approach to validity offers a complementary and important addition to the range of validation approaches.

3.5 The Structuralist Approach

If, roughly cast, the modeler's form of validity focuses on theory fitting data, and the philosopher's form of validity focuses on whether theory is well-grounded in an epistemological and ontological conceptual framework, then the structuralist form of validity can be viewed as iterating between both these forms [41].

As a case in point, consider Piaget's cognitive-developmental theory wherein he sought to characterize the development of the structure of human knowledge. That ambitious undertaking led Piaget to posit four overarching stages in development from infancy through late adolescence – sensory motor, preoperations, concrete operations, and formal operations – through which Piaget characterized how children's thinking increased in its capacity for synthesis and abstraction. Characterizations of each stage are comprised of detailed formal properties. For example, one central property of concrete operations is that of reversibility, that in the movement from one conception to another, the original conception can be simultaneously invoked.

How, then, did Piaget seek to establish the validity of his stage characterizations? In part, he drew on philosophical theory of what counted as domains of knowledge – such as of physics, space, time, logic, and number. For each of these domains, he further drew on philosophical criteria of what comprised those domains; and then he sought to find their origins in childhood. Thus, for example, Piaget drew widely from mathematicians such as Russell and Whitehead to help formulate the ideas for logical groupings [43]. But, for Piaget, the philosophical criteria by themselves did not establish validity. Rather the theory led to investigations of how the child formed concepts about these domains in the world, which in turn led to revisions in the theory of what those domains consisted of, which in turn led to further refined investigations, more empirical data, and so it went, back and forth between theory and data, for the entire span of Piaget's long career. Piaget called this approach *genetic epistemology* [42]. By this term, he meant that valid knowledge did not exist "out there" to be discovered in a world independent of humans, as proposed by correspondence theories of epistemology, nor is valid knowledge just socially constructed, as in it being anything we want it to be or might imagine it to be. Rather valid knowledge is constructed through interaction with the reality of an external world and the biological affordances and constraints of our species.

Our approach to establishing the validity of our interaction patterns can often be best understood by this structuralist's approach. The interaction pattern of the Introduction, for example, draws substance from unpacking the form and function of Introductions across time, context, and culture; but it further needs empirical evidence to increase its validity; and in the process of gaining such evidence, more conceptual understanding of the pattern itself becomes revealed – and the iterative process is underway.

One difference in Piaget's structuralist approach and ours is that Piaget focused on the structure of individual cognition, and its ontogenesis. In contrast, with our focus here on interaction patterns, we are interested in the structure of human interaction between the organism and its environment. Our endeavor with interaction patterns is more akin to ecological psychologists like Gibson [22, 23] who focused on person-environment interaction as the unit of analysis; though we differ from many ecological psychologists in not positing a pragmatist epistemology, where truth is what "works" [27] – but a broader epistemology that is amenable to the different forms of validity distinguished in this paper.

4 NEXT STEPS

In our view, all of the above forms of validity can apply to interaction patterns, though perhaps the structuralist approach is the most relevant. What, then, might next steps look like in generating and validating interaction patterns? First off, we need to build up a large set of possible interaction patterns, perhaps in the range of 50-70. Why not 15 or 250? We need enough interaction patterns to begin to map out the diversity of interactions, so as to make this endeavor compelling. Fifteen is too few. Two hundred and fifty seems too ambitious at this early stage.

In generating this first large group of interaction patterns, why will we choose some over others? Our answer is that some of our patterns – such as the Introduction – seem to us fundamental for motivating many forms of human-robot interaction, and thus warrant particular attention. Other patterns tap into what will become huge societal issues (and concerns) in the immediate future, and thus warrant our particular attention. For example, the interaction pattern of physical intimacy with a robot has been to date sensationalized by such books as *Love and Sex with Robots* [36]. By focusing on less sexual instantiations of physical intimacy, our immediate work will let us speak to these issues on a societal level, and shape healthy directions for this interaction pattern in HRI. Still other patterns (such as physical motion) are opportunistic in that we are building on the existing repertoire that is currently engineered autonomously in the robot we have been working with (ATR's "Robovie"). Finally other patterns will be chosen to fit different categories of interaction, per our theoretical commitments of how to understand the ontological nature of human-robot interaction: that is, it can be like human-human interaction, human-animal interaction, and human-machine interaction, as well as something uniquely itself [32].

Two concerns in particular could be raised about our general approach to interaction patterns. One concern is this: Are there an infinite number of interaction patterns that correspond to the infinite number of ways humans interact with other humans, animals, nature, and artifacts? If so, there's no end to this project. Our answer is that an endeavor like this one might seem to the reader as endlessly open-ended, as one can seemingly always come up with a new interaction pattern in HRI. But at some point, "new" patterns will be seen as providing very little new information to an existing pattern, and thus either subsumed within an existing pattern or recognized as an instantiation of an existing pattern. The validity claim here is somewhat akin to that of parsimony by the modeler (Section 3.3).

A second concern is that we have been speaking of the interaction patterns as independent entities. But is it not the case that many patterns interact with one another, and do not exist in discrete units? Our answer is that our laboratory work in part does aim first to carve out discrete units, but eventually to link and integrate them. By analogy, Alexander articulated his 253 patterns as discrete units but emphasized how they can be combined into a "pattern language." Framed as our long-term goals, we aim for something similar: an interaction pattern language for sociality in HRI.

5 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. Psychometrics is just one, though often the most referenced in the social-scientific communities. Thus, in this paper, we have offered five overarching approaches to validation. To portray these approaches, we used interaction patterns as a case in point. As the HRI field evolves, many useful approaches for characterizing social interaction between humans and robots will emerge. Thus the approaches toward establishing validity as sketched in this paper may well have broad appeal.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant Numbers IIS-0842832 and IIS 0905289. 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] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 'Technical recommendations for psychological tests and diagnostic techniques', *Psychological Bulletin,* **51,** supplement, 1954.

[2] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, *Standards for Educational and Psychological Testing,* American Educational Research Association, Washington, D.C., 1966.

[3] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, *Standards for Educational and Psychological Testing,* American Educational Research Association, Washington, D.C., 1999.

[4] C. Alexander, S. Ishikawa, and M. Silverstein, *A Pattern Language: Towns, Buildings, Construction,* Oxford University Press, New York, 1977.

[5] D. Bernstein, K. Crowley and I. Nourbakhsh, 'Working with a robot: Exploring relationship potential in human–robot systems', *Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems*, **8,** 465-482, (2007).

[6] C.L. Breazeal, *Designing Sociable Robots: Intelligent Robots and Autonomous Agents*, MIT Press, Cambridge, MA, 2002.

[7] J. Brochers, *A Pattern Approach to Interaction Design,* John Wiley and Sons, Chichester, England, 2001.

[8] G. Casaday, *Notes on a Pattern Language for Interactive Usability,* Paper presented at the CHI '97 Extended Abstracts on Human Factors in Computing Systems, Atlanta, Georgia, (1997).

[9] J. Cassell and A. Tartaro, 'Intersubjectivity in human–agent interaction', *Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems*, **8**, 391-410, (2007).

[10] E.S. Chung, J.I. Hong, J. Lin, M.K. Prabaker, J.A. Landay, and A.L. Liu, 'Development and evaluation of emerging design patterns for ubiquitious computing', in *Proceedings of the 2004 Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques,* Cambridge, MA, USA, 233-242, (2004).

[11] V. Corkum and C. Moore, 'The origins of joint visual attention in infants', *Developmental Psychology,* **34**, 28-38 (1998).

[12] L.J. Cronbach, 'Test validation', in *Educational Measurement*, ed., R.L. Thorndike, American Council on Education, Washington, D.C., 2nd edn., 443-507, (1971).

[13] K. Dautenhan and I. Werry, 'Towards interactive robots in autism therapy', *Pragmatics and Cognition*, **12**, 1-35, (2004).

[14] H. Dingle, *Science and literay criticism,* Nelson, London, 1949.

[15] D. Feil-Seifer, K. Skinner and M.J. Matarić. 'Benchmarks for evaluating socially assistive robotics', *Interaction Studies: Social*

Behavior and Communication in Biological and Artificial Systems, 8, 423-439, (2007).

[16] S.E. Fish, *Is There a Text in This Class?: The Authority of Interpretive Communities,* Harvard University Press, Cambridge, MA, 1980.

[17] T. Fong, I. Nourbakhsh, and K. Dautenhahn, 'A survey of socially interactive robots', *Robotics and Autonomous Systems*, **42**, 143-166, (2003).

[18] R.P. Gabriel, *Patterns of Software: Tales from the Software Community,* Oxford University Press, New York, 1996.

[19] H.G. Gadamer, *Philosophical Hermeneutics,* trans. D.E. Linge, University of California Press, Berkeley, 1977.

[20] H. Gardner, *Frames of mind: The theory of multiple intelligences.* Basic Books, New York, 1983.

[21] K.J. Gergen, *An Invitation to Social Construction,* Sage, London, 1999.

[22] J.J. Gibson, *The Senses Considered as Perceptual Systems,* Houghton-Mifflin, Boston, 1966.

[23] J.J. Gibson, *The Ecological Approach to Visual Perception,* Lawerence Erlbaum, Hillsdale, NJ, 1986 (original work published 1979).

[24] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, et al., 'Designing robots for long-term social interaction, in *Proceedings of IROS,* IEEE, Picataway, NJ, 305-312, (2005).

[25] I. Graham, *A Pattern Language for Web Usability,* Addison-Wesley, London, England, 2003.

[26] J. Habermas, *Communication and the Evolution of Society,* trans., T. McCarthy, Beacon Press, Boston, MA, 1979.

[27] H. Heft, 'The social constituion of perceiver-environment reciprocity, *Ecological Psychology,* **19**, 85-105, (2007).

[28] E.D. Hirsch, *Validity in Interpretation*, Yale University Press, New Haven, 1967.

[29] M.A. Honey, 'The interview as text: Hermeneutics considered as a model for analyzing the clincially informed research interview, *Human Development,* **30,** 69-82, (1987).

[30] P.H. Kahn, Jr., 'Children's obligatory and discretionary moral judgments, *Child Development*, **63**, 416-430, (1992).

[31] 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* (HRI '08), Amsterdam, Netherlands, 97-104, (2008).

[32] P.H. Kahn, Jr., B. Friedman, D.R. Pérez-Granados, and N.G. Freier, 'Robotic pets in the lives of preschool children', *Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems*, **7**, 405-436, (2006).

[33] P.H. Kahn, Jr., H. Ishiguro, B. Friedman, T. Kanda, N.G. Freier, R.L. Severson, and J. Miller, 'What is human?- Toward psychological benchmarks in the field of human-robot interaction', *Interaction Studies: Social Behavior and Communication in Biological and Artifical Systems,* **8**, 363-390, (2007).

[34] T. Kanda, T. Hirano, D. Eaton and H. Ishiguro, 'Interactive robots as social partners and peer tutors for children: A field trial', *Human-Computer Interaction*, **19**, 61-84, (2004).

[35] M. Killen, 'Children's social and moral reasoning about exclusion, *Current Directions in Psychological Science,* **16**, 32-26, (2007).

[36] D. Levy, *Love and Sex with Robots. The Evoluton of Human-Robot Relationships,* Harper Collins, New York, 2007.

[37] K.F. MacDorman and P.H. Kahn, Jr., 'Introduction to the special issues on psychological benchmarks of human-robot interaction', *Interaction Studies: Social Behavior and Communication in Biological and Artifical Systems,* **8**, 359-362, (2007).

[38] S. McNamee and K.J. Gergen, 'Therapy as social construction', *Inquiries in Social Construction,* Sage, London, (1992).

[39] S. Messick, 'Validity', in *Educational Measurement,* ed., R.L. Linn, MacMillan, New York, 3rd edn., 13-103, (1989).

[40] S. Messick, 'Validity of psychological assessment', *American Psychologist,* **50**, 741-749, (1995).

[41] J. Piaget, *Structuralism,* Harper and Row, New York, 1970.

[42] J. Piaget, *Genetic epistemology*. W.W. Norton & Co., New York, 1971.

[43] J. Piaget, 'Piaget's theory', in *Handbook of Child Psychology: Vol 1. History, Theory, and Methods,* ed, W. Kessen, series ed, P.H. Mussen, 4th edn., 103-128, Wiley, New York, (1983).

[44] J. Rawls, A *Theory of Justice*, Harvard University Press, Cambridge, 1971.

[45] N.A. Salingaros, 'The structure of pattern languages', *Architectural Research Quarterly,* **4**, 149-161, (2000).

[46] M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita, 'Interactive humanoid robots for a science museum, in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction* (HRI '06), Salt Lake City, Utah, USA, 305-312, (2006).

[47] C.S. Taber and R.J. Timpone, R. J, *Computational modeling*. Sage university papers series, no. 07-113, Sage Publications, Thousand Oaks, Calif, (1996).

[48] F. Tanaka, A. Cicourel, and J.R. Movellan, 'Socialization between toddlers and robots at an early childhood education center', *Proceedings of the National Academy of Science,* **104**, 17954-17958, (2007).

[49] J.G. Trafton, A.M. Harrison, B. Fransen, and M. Bugajska, 'An embodied model of infant gaze-following', *International Conference of Cognitive Modeling,* (2009).

[50] E. Turiel, 'The development of morality', in *Handbook of child psychology. Social, emotional, and personality development,* eds W. Damon and N. Eisenberg, J. Wiley, New York, NY, 863-932, 5th edn, (1998).

[51] S. Turkle, 'Authenticity in the age of digital companions', Interaction Studies: Social Behavior and Communication in Biological and Artifical Systems, **8**, 501-517, (2007).