



ELSEVIER

Contents lists available at SciVerse ScienceDirect

Language & Communication

journal homepage: www.elsevier.com/locate/langcom

Huckleberry Finn takes the Turing test: The transformation of ontologies and the virtuality of kinds



Paul Kockelman

Department of Anthropology, Barnard College, Columbia University, 3009 Broadway, New York, NY 10027, United States

ARTICLE INFO

Article history:

Keywords:

Turing Test
Ontology
Inference
Identity
Virtuality
Performativity

ABSTRACT

This essay is about the relation between social statuses, mental states, and material substances; the indexical signs used to infer such underlying kinds; the conditions for and consequences of the ontologies that license such inferences; and the potentially reflexive and transformative relations individuals bearing such identities have towards each other and themselves. While it begins with what may be called ‘the Huckleberry Finn Test’ (inferring gender in face-to-face interaction), it concludes with the Turing Test (deciding between human and computer in teletype-mediated communication). It argues that most thought about the Turing test has focused on a very limited type of inference. And it shows four other important ways our indexical encounters with others can both transform, and be transformed by, our ontologies.

© 2013 Elsevier Ltd. All rights reserved.

And don't go about women in that old calico. You do a girl tolerable poor, but you might fool men, maybe. Bless you, child, when you set out to thread a needle don't hold the thread still and fetch the needle up to it; hold the needle still and poke the thread at it; that's the way a woman most always does, but a man always does t'other way. And when you throw at a rat or anything, hitch yourself up a tiptoe and fetch your hand up over your head as awkward as you can, and miss your rat about six or seven foot. Throw stiff-armed from the shoulder, like there was a pivot there for it to turn on, like a girl; not from the wrist and elbow, with your arm out to one side, like a boy. And, mind you, when a girl tries to catch anything in her lap she throws her knees apart; she don't clap them together, the way you did when you caught the lump of lead. Why, I spotted you for a boy when you was threading the needle; and I contrived the other things just to make certain.

—Adventures of Huckleberry Finn, Chapter 11, by Mark Twain (2006 [1884])

1. All kinds are virtual

Dressed as a girl, Huckleberry Finn went into town to find out what people were saying about Jim. In this scene, Mrs. Judith Loftus has just ‘spotted him for a boy’, and she is reporting to him the evidence she used to come to this conclusion. In particular, she has a set of assumptions regarding the types of behaviors that boys and girls are more or less likely to do in various circumstances. For example, girls not only wear dresses and bonnets when they are in public, but they also open their knees when trying to catch something in their lap, and hold their arm stiffly when throwing. More specifically, Mrs. Loftus has a relatively elaborate (and, in part, articulatable) set of assumptions about which indices are evinced by individuals belonging to what kinds with what likelihoods (e.g. ‘most always’). And she uses these assumptions not only to infer that Huck is a boy rather than a girl, but also to generate a set of experiments (or ‘trials’) to check her own hypothesis.

E-mail address: pk2113@columbia.edu

Some of these indices are relatively easy to think of and simple to feign (e.g. wearing a bonnet). Others are relatively tacit and embodied, and so hard to predict or hide (e.g. threading a needle). And all are strongly correlated with one social kind rather than another (if only in the mind of Mrs. Loftus), and involve circumstances and behaviors that are more or less public and easy to elicit.

Such a set of assumptions might be called a theory (when articulated in relation to a scientific institution), a stereotype (when negatively valenced), a likelihood (when framed mathematically), a heuristic (when framed qualitatively, or as a 'rule of thumb'), an imaginary (when understood in relation to an underlying account or narrative about the prototypic entities involved in the domain being judged), a culture (when more or less intersubjectively shared by a group of people), and even a habitus or 'sense' (when understood as a tacit intuition regarding another's identity via their techniques of the body, styles of speaking, and so forth). In what follows, the term *ontology* will be used as a cover-all term to capture the ramifications present in each of these framings (Kockelman, 2013). Indeed, just as Huck had an ontology regarding which indices boys and girls were likely to express (even if he was out-ontologized by his host), and just as Mrs. Loftus had an ontology regarding which indices would be hard to mask, so Mark Twain had an ontology regarding which ontologies individuals with identities like Huck and Mrs. Loftus would be likely to hold.

This essay is about the relation between such signs and the kinds they index; the conditions for and consequences of the ontologies (theories, likelihoods, imaginaries, etc.) that license such inferences; and the potentially reflexive (or self-conscious) and destabilizing (or ontology-transforming) relations individuals bearing such identities have towards each other and themselves. In addition to social statuses (like boy and girl), it also takes into account mental states (like belief versus doubt) and material substances (like human versus machine). It thereby focuses on the relations between social-cultural, cognitive-affective, and material-technical modes of identity; the types of signs that inferentially and indexically lead to and follow from such underlying kinds; and the epistemic, affective, and moral commitments such semiotic processes involve. And though it is grounded in the dynamics of real-time, face-to-face interaction, it develops the repercussions of its analysis for digitally mediated and pervasively networked forms of interaction. While it thus began with what may be called 'the Huckleberry Finn Test' (boy versus girl in face-to-face interaction), it will conclude with a discussion of the Turing Test (computer versus human in teletype-mediated communication). It argues that most thought about this test has focused on a very limited type of inference. And it shows four other important ways our indexical encounters with others can both transform, and be transformed by, our ontologies.

2. Ontologies and their transformations

As introduced above, the term *index* will be used to refer to any quality that is relatively perceivable to some agent (e.g. actions like wearing a bonnet and threading needles). The term *kind* will be used to refer to any projected propensity to exhibit particular indices (e.g. boy and girl). The term *agent* will be used to refer to any entity that can perceive such an index and thereby project such a kind (e.g. Mrs. Loftus). The term *individual* will be used to refer to any entity that can evince indices to an agent and thereby be a site to project kindedness by that agent (e.g. Huckleberry Finn). And the term *ontology* will be used to refer to an agent's assumptions as to the indices, kinds, and individuals that constitute a particular world (e.g. the partially articulatable beliefs of Mrs. Loftus). See Table 1.

Crucially, not only are social statuses (speaker, banker, woman, etc.) kinds, but so are mental states (believing X, fearing Y, etc.), and material substances (gold, water, snow, etc.). In particular, interpreting agents can project such kinds onto particular individuals (such as *this stuff*, *that woman*, *my dog*) as a function of the indices they evince (the clothes they wear, the actions they undertake, the temperatures at which they freeze, the things they say, the relations they have to each other, and so forth). *That's gold, she's a banker, he's afraid of the dark*. In this way, ontologies drive interpretation: by your index (sign), I may infer your kind (object), and thereby come to expect (interpretant) other indices that would be in keeping with your kind (insofar as I have a particular ontology).

But rather than focusing on how *ontologies mediate interpretations*, which is as far as we can get with a single passage from Huckleberry Finn, we are also interested in how *interpretations mediate ontologies*.

For present purposes, there are five kinds of ontological transformativity—whereby an interpreting agent's ontology transforms via indexical encounters with an individual. See Table 2. The first kind of transformativity is simply performativity in a relatively generalized sense: some index, or ensemble of indices, may change an individual's kind more or less irrespective of some particular agent's assumptions about it. Here go all the usual processes that produce kinds in the first place, from chemical reactions to marriage ceremonies, from performative utterances to contractual agreements, from socialization practices to evolutionary processes. The second kind of transformativity is perhaps the most quotidian and is often relatively

Table 1
Some key constituents of ontologies.

<i>Index</i>	Any quality that is relatively perceivable (to some agent)
<i>Kind</i>	Any projected propensity to exhibit particular indices
<i>Agent</i>	Any entity that can perceive such an index and project such a kind (itself often an individual)
<i>Individual</i>	Any entity that can evince indices (to an agent) and thereby be a site to project kindedness (by that agent)
<i>Ontology</i>	The assumptions an agent has as to the indices, kinds, and individuals that constitute a particular world

Table 2
Some kinds of ontological transformativity.

-
- (1) Indices (and signs more generally) may change an individual's kind irrespective of an agent's ontological assumptions
 - (2) Indices may change an agent's ontological assumptions regarding the kinds that constitute a particular individual
 - (3) Indices may change an agent's ontological assumptions regarding the indices that constitute a particular kind
 - (4) Indices may change an agent's ontological assumptions regarding the indices, individuals, kinds, and agents that constitute a particular world
 - (5) Changes in an agent's ontological assumptions about a world (in any of the foregoing ways) may change the world about which the agent makes assumptions
-

deductive: indices may change an agent's ontological assumptions regarding the kinds that constitute a particular individual. For example, from your apron, I infer you are a busboy rather than a waiter. This is where Mrs. Loftus aimed her inquiry, as well as legions of theorists about the Turing Test. The third kind of transformativity is often relatively inductive: indices may change an agent's ontological assumptions regarding the indices that constitute a particular kind. For example, from your behavior, I infer that waiters sometimes bus dishes. Had Mrs. Loftus, in her encounter with Huck, changed her assumptions about the throwing and catching abilities of boys and girls, this kind of transformativity would have been operative. The fourth kind of transformativity is often relatively abductive: Indices may change an agent's ontological assumptions regarding the indices, individuals, kinds, and agents that constitute a particular world (as well their assumptions regarding the possibilities of other worlds that could be constituted). For example, from your behavior, I hypothesize a new social status (say, the maitre d'). Had Mrs. Loftus hypothesized a new status—say, the transvestite—this kind of transformativity would have been operative. And finally, there is a fifth kind of transformativity that may involve any of the other kinds: in particular, my assumptions about the world (as to its individuals, indices, and kinds) may transform the world about which I make assumptions. If Huck internalized part of Mrs. Loftus's ontology, and so came to act more (or less) in line with her assumptions, or came to raise his own daughter or son to throw and catch in a new manner, or came to disguise himself differently in the future, this kind of transformativity would be operative.¹

3. The Turing test

We may now switch to the famous scenario devised by Turing (2004 [1950]). Imagine you are alone in a room, with a keyboard and monitor, having a text-based conversation, via teletype, with someone (or something) who claims to be a woman, but whose real identity is unknown to you. Indeed, not only may it be a man rather than a woman, but it may also be a machine rather than a human. You can type in whatever you want, and this someone else will reply; and vice versa. And thus the only evidence you can gather as to its kindness is text-based: what kinds of questions does it ask you (with what kinds of syntax, semantics, and pragmatics); and how does it answer your questions. In some sense, then, you are playing the role of Mrs. Loftus trying to ferret out the identity of a suspicious guest.

What text-based patterns might you look for, or text-based experiments might you run, to determine the identity of this other? If you can or cannot determine the identity of this other via such trials, what does this say, philosophically, about the relation between humans and machines (or men and women)? What does this say, technically, about the relation between computers and their software (or actors and their techniques)? And, more generally, what does it tell us about the potentially unsettling effects of computer-mediated communication in relation to face-to-face interaction?

My aim in what follows is not to answer these questions per se. (Indeed, merely summarizing the rich secondary literature on the Turing Test, and hence over fifty years of thought about this topic, is a book in itself.²) Rather, I want to use Turing's beautifully imagined scenario to test the utility of the foregoing analytic framework—itsself grounded in a particular theory of the relation between ontology, transformativity, and virtuality. Indeed, given the similarities between this scenario and the situation in which Huck and Mrs. Loftus found themselves, everything said above is immediately applicable—indeed, it was

¹ While there is much more that could be said about this framework (see Kockelman 2013 for the details), including a huge host of qualifications and caveats, the bare bones formulation undertaken in this section is enough for what follows. That said, it should be stressed that all the units are relationally defined (e.g. an agent is being defined in terms of its ability to sense indices and project kinds, such that what counts as an index is dependent on what can be sensed by an agent, and so on). It should be stressed that agents and individuals need not be 'individuals' (qua specific people), but may be people, things, or anything outside or in-between (parts, collectivities, relations, and so on). It should be stressed that ontologies are recursive and frame-dependent: for example, an index (of one kind) may itself be a kind; and having a particular ontology may itself be a complicated index (of another kind). And it should be stressed that the claim here is not that there are three kinds of entities in the world (social statuses, material substances, and mental states), but that three entities that are often, in one widespread social-science imaginary, treated in different terms (e.g. normatively regimented versus causally regimented; or representational versus non-representational), may be theorized in similar terms for the sake of semiotic-inferential processes of identification (whether or not they exist, howsoever they may be imagined or regimented, and whatever their actual nature). Other particular relevant kinds, for example, include textual genres (e.g. spam versus ham) and singular identities (e.g. John versus Jake). Kockelman (submitted for publication) takes up the mathematics of the inferences involved, paying particular attention to spam filters and sieving devices more generally.

² See, for example, French (1990) and Saygin et al. (2001). Crucially, Turing himself (2004 [1950], p. 453) comes closest to foregrounding the cultural specificity of ontologies, and the kinds of inductions that lead to them, and the kinds of false (or transient) conclusions that can be drawn from them (under the heading 'Arguments from Various Disabilities'). In this way, he noted the importances of issues underlying transformativity #3, even though he, like most subsequent scholarship, spent most of his time on transformativity #2.

designed precisely to deal with such a scenario. This section, then, will serve as much a summary, as an extension, of my arguments.

For the moment (and quite naively as will be seen below) we may focus on the first issue—the text-based trials one might run to infer the true identity of this other. And we may take as our basic premise the following claim: *it should be able to do to us whatever we are doing to it* (or at least a significant chunk of this). Phrased another way, the key indices for us to look for as signs of its true kind (human versus machine) would turn on its ability to track our indices, as well as on its ability to track our tracking of its indices—not perfectly of course, but to some degree, which is certainly one key feature of the human-kind (at least in Goffman's (1959) ontology). More carefully put, as per section 2, from the indices we have evinced, it should be able to make inferences about our mental states, social statuses and material substances; and from such inferred mental states, social statuses and material substances, it should be able to expect other indices we may evince (and sanction us in the light of those expectations, or draw other inferences from our failures to meet those expectations). In short, it should be able to project kinds onto individuals through their indices via its ontology. In effect, then, we would need to track indices of its ability (qua projected propensity) to track indices of our propensities.

But, of course, this is just one possibility. And while it might be necessary to pass the test, it's surely not sufficient. And so it will forever be tempting to imagine a range of other abilities that would provide necessary and sufficient criteria—and thereby secure some criterial definition of the authentically human. To take just one example, which we can call the singular-humanist criterion, let us suppose that the unknown other should have also an 'individual identity' in some metaphysical sense. This might mean leaving a biographical trace in the mental states, social statuses and material substances of others (a notion closely related to Benjamin's concept of aura). It might mean having a standard of values which gives its mental states and social statuses a kind of coherence (as per the arguments of Charles Taylor or Max Weber). It might mean that it has a kind of daimon, soul, or telos that gives its travels through the world a kind of biographical closure (loosely following some ideas of Hannah Arendt). It might require that the entity have a reflexive self of the Jamesean sort—such that not only does it have qualities (such as social statuses, mental states and material substances, or a ensemble of belongings more generally), but that the flourishing or foundering of such qualities matter to it. Or, indeed, it might mean that its behavior, or 'life', is subject to rich and multiple interpretations (like any great character in literature or history—such as Huckleberry Finn or Alan Turing). And so on, and so forth. However such a unique identity be framed, perhaps the crucial point here would be that to be human (as a type) is to be both cultural and biographical, or singular with respect to both the community to which one belongs and the life-path along which one travels. In short, it might be argued that a human being without culture and personality (and thus an historically and personally singular ensemble of social statuses, mental states and material substances) is not human at all.³

We may now turn to the second caveat, which may be understood as a counter to precisely such naive modes of speculation (regarding putative human-specific propensities, such as ontological projection and singular humanism). In particular, the Turing Test is perhaps best understood not as a benchmark for artificial intelligence, but as a diagnostic of what a given group of people at a given point in time think is the essence of the human (or the limitations of the machine)—given the prejudices (or progress) of their era, discipline, expertise, culture, and so on. That is, the Turing Test is really a kind of Rorschach test for the questioner's sense of self, current theories of the brain and body relation, the robustness of our knowledge of a population's statistical profile, and more or less fashionable ideas about putative human-specific processes (e.g. recursion, meta-cognition, sub-cognition, joint attention, theory of mind, performativity, intersubjectivity, shared intentionality, singular-humanism, and so forth). In this way, the real value of the Turing Test is akin to science fiction: functioning not as an augury of the possibilities to come, but as a symptom of the prejudices that are. This very essay, then, is precisely an assay: an attempt to make maximally public and minimally ambiguous the author's (and/or his discipline's) own personal and cultural prejudices (as to how we may know what kinds of things there might be).

Ironically, such prejudices are essentially what we called ontologies (theories, stereotypes, heuristics, intuitions, likelihoods or imaginaries) in section 1. Indeed, somewhat damningly, the foregoing discussion of the Turing Test, and almost all the literature written on it, presumes that transformativity #2 is the central question. That is, given some indices (i.e. actual teletype-based performances of the other, however richly imagined or creatively theorized), let us try to infer the kind-ness of our unknown interlocutor. And so there has been no end of discussion as to what constitutes the truly human (as a real kind), and speculation as to how we might be able to program a computer to mimic it (as a virtual kind). Only partially tongue-in-cheek, one is tempted to say that the crucial index of individuals belonging to the human kind is their so-often-indexed propensity to articulate their own ontological assumptions as to what are the crucial indices of the essential propensities of the human kind. . . .

Crucially, however, the most interesting aspect of the Turing Test may not be transformativity #2, but rather transformativities #3 and #4. For example, the point of such an encounter with another might be to generate new ontologies for old kinds; or, even better, to generate hypotheses regarding the existence of new kinds, or regarding the existence of hitherto unknown indices (for old kinds). That is, the real dialog might not be between me and the unknown other, but rather between 'us' (as an epistemic community, intent on answering such questions) before and after we have engaged in such a dialog—such that we learn more about what both things like it and people like us are capable of becoming. Such a stance would represent the marriage of Alan Turing's test with Max Weber's ideal type (1949), and thus the marriage of mechanized

³ See, for example, Arendt (1958), Benjamin (1968), James (1950), Taylor (1989), and Weber (1978).

circuits and hermeneutic circles: a way to goad humans (and machines) into trying out ever new possibilities for being-human in relation to being-machine.

Finally, there is still that pesky transformativity #5, itself directly linked to transformativity #1: the possibility that the other may internalize our assumptions about it, and thus come to act according to our ontologies; thereby confirming, rather than contradicting—and possibly transforming—our assumptions about it. Indeed, for some, this last scenario may be the most emblematic index that machines are capable of human-being, and perhaps 'kind enough' to *be* human: when we internalize our machines' ontologies about us, and thereby come to act according to their assumptions about our behavior (and vice versa).

Acknowledgements

I want to thank Paul Manning, Talbot J. Taylor, and especially Nick Enfield for very helpful comments and encouragement.

References

- Arendt, Hannah, 1958. *The Human Condition*. University of Chicago Press, Chicago.
- Benjamin, Walter, 1968. 'The Work of Art in the Age of Mechanical Reproduction', in *Illuminations*. Schocken Books, New York, pp. 217–252.
- French, Robert.M., 1990. The Turing test: the first 50 years. *Trends in Cognitive Science* 4 (3), 115–122.
- Goffman, Erving., 1959. *The Presentation of Self in Everyday Life*. Anchor Books, New York.
- James, William, 1950. *The Principles of Psychology*. Dover Publications, New York.
- Kockelman, Paul, 2013. *Agent, Person, Subject, Self: A Theory of Ontology, Interaction, and Infrastructure*. Oxford University Press, Oxford.
- Kockelman, Paul, submitted for publication. *Hunting Ham and Sieving Spam: Ontological Transformativity and Bayesian Anthropology*.
- Saygin, Ayse Pinar, Cicekli, Ilyas, Akman, Varol, 2001. Turing test: 50 years later. *Minds and Machines* 10, 463–518.
- Taylor, Charles, 1989. *Sources of the Self*. Harvard University Press, Cambridge.
- Turing, A.M., 2004 [1950]. In: Copeland, B. Jack (Ed.), 'Computing Machinery and Intelligence' in *The Essential Turing*. Oxford University Press, Oxford, pp. 441–464.
- Twain, Mark, 2006 [1884]. *The Adventures of Huckleberry Finn*. Project Gutenberg Ebook.
- Weber, Max, 1949. *The Methodology of the Social Sciences*. Free Press, New York.