

Science Considered Harmful

Ian Horswill

Northwestern University, 2133 Sheridan Road, Evanston IL 60208
ian@northwestern.edu

Abstract

As the INT community matures, one hears increasing calls for the field to become more scientific, and even prior restraint in which researchers avoid a project entirely out fear it wouldn't be sufficiently scientific. In this position paper, I'll argue that science, and more generally rigor, are not well served by a discourse that labels things as "science" and "not science." In particular, I will argue that we need to open ourselves to the adoption of methods from the arts, including case studies and the practice of studio critique.

Introduction

Most members of the INT community, myself included, are trained as computer scientists. Natural inclination and institutional pressures direct us toward the methodologies of good engineering research, often referred to colloquially as doing "science" in contrast to mere "engineering." As the field has matured there have understandably been calls for us to be more scientific. Indeed at a panel at INT4, Mark Riedl argued that the time might be right for a "science of aesthetics" (Horswill, Riedl, & Young, 2011). However, a science of narrative means different things to different people; Todorov (1969) thought narratology already *was* the science of narrative, while members of the INT community presumably want something with more user studies and PDDL code.

In the more than 2000 years since Aristotle's *Poetics*, there have been periodic attempts to explain aesthetic domains using what one might refer to as scientific methodologies; the most successful of these being perhaps the attempt to explain the traditional Western tonal theory of in terms of harmonic (i.e. Fourier) structure, and the use of Gestalt theory in 20th century Western graphic design. And yet, with a few such exceptions, these efforts have had virtually no influence on contemporary art theory, practice, criticism, or education. While one might attempt to

explain this in terms of some psychological or political resistance on the part of artists and critics, one is nevertheless hard-pressed to find scientific theories that are particularly useful either in either making or critiquing art works. If Fourier theory were powerful enough to predict what was a good melody, then algorithmic composition would have become a force in popular music 40 years ago. Lexical-functional grammar, while a powerful theory of language, tells us nothing about poetics. And the theory of visual attention, while interesting and powerful, doesn't offer Hitchcock any lessons on how to direct a viewer's attention in a scene that he didn't already know.

So while it's not impossible that a group of AI people might develop a science of narrative that was sufficiently powerful to be able to explain and predict what made a good story, one has to admit that it's a bold claim.

What would a science of narrative even mean?

The key issue is that the terms *science* and *narrative* each hide complex systems of ambiguities. They can be scoped almost arbitrarily broadly or narrowly, and claims that are perfectly reasonable under one reading of the terms may be blatantly false under another. When we forget these ambiguities and use the terms as if they were rigid designators we can get into trouble.

What do we mean by narrative?

Narrative covers a lot of ground: newspaper stories, criminal confessions, interactive fiction, *The Sims*, most films, some music, radio plays, textbook descriptions of The Big Bang, Broadway musical theatre, and much poetry. It even includes novels. While many of these have common characteristics such as character development, dramatic tension, and Aristotelian arcs, many do not. If we search for a set of characteristics shared by all, it's unclear we can do better than "set of (mostly) causally related events." To the extent that most authors mean something more when they use the term, but not necessarily the "same

more” as when other authors use the term, it may be more useful to think of narrative as a radial category (Lakoff, 1990) or Wittgensteinian family resemblance (Wittgenstein, 1958). That is, we might be better off abandoning the notion that narrative was a single, clearly defined category (a natural kind).

To see why we might get ourselves into trouble by thinking of narrative as being a natural kind, consider the following claims, which most of us would believe:

1. Narrative is important to human cognition.
2. Humans prefer narratives that contain conflict.

Indeed, it’s common to start with propositions like these at the beginning of talks in our community.¹ The problem is that the two claims involve incompatible senses of the term narrative. Sense (2) involves people’s *taste* in stories. Prototypical cases of narrative in that sense would be novels, films, and so on. Sense (1), however, is much broader, meaning something along the lines of information about sequences of causally related events in time, generally involving agents with goals and intentions. Prototypical cases of this sense of the term would be frames and schemata such as Schank and Abelson’s “scripts” (1977).

Let’s unpack the term narrative in each claim:²

1. *Representations of causal sequences involving agents and goals* are important to human cognition.
2. Humans prefer *literary narratives* that contain conflict.

These claims are uncontroversial, but if we swap the unpackings, the results are much more dubious:

- * *Literary narratives* are important to human cognition.
- * Humans prefer *representations of causal sequences involving agents and goals* that contain conflict.

This conflation of the broad and narrow senses of narrative dates back at least to Barthes (1966), continues in modern narratology (Abbott, 2008), and has been reinvented by the AI community.

The conflation matters because it’s hiding a deep theoretical tension in the meaning of narrative, one that calls into question whether it’s even possible to have a single science of narrative. The different senses of narrative may well require different methodologies.

¹ To be precise, it’s common to begin with the observation that narrative has some privileged role in human cognition and then to pivot to the issue of what makes a good story.

² These are not particularly good glosses of the term narrative; you may prefer others. The point is that no one gloss works well for both claims.

What do we mean by science?

Dictionary.com shows 5 different senses of the term *science*, ranging from “natural sciences,” which is so narrow as to rule out narrative, to “knowledge, as of facts or principles,” which is so broad as to include nearly everything done in academia. None of the definitions given are useful for distinguishing a specifically *scientific* theory of narrative from an unscientific one.

Eagleton begins his classic text on literary theory (Eagleton, 2008) with the seemingly trivial question “what is literature,” only to show that every candidate definition is at odds with the way the term is used in real life. In the end, he concludes that the only definition that matches empirical usage is that literature is *highly valued writing*. One therefore cannot examine certain issues in literature without examining issues of value, and therefore culture, class, and so on.

So far as I can tell, computer scientists use *science* in the same way: science is *highly valued methodology*. I’ll offer three pieces of evidence for this. First, there is already such diversity in the accepted methods of computer science, that it’s hard to find any core: the methods of a complexity theorist and an HCI researcher are nearly disjoint. Second, when I’ve asked colleagues who declare, “That’s not science”³ to unpack what they meant by the statement, they give widely varying answers (all real examples):

- That’s a program, not a theory
- That doesn’t have data
- That *merely* has data, it doesn’t have theorems
- You can’t get NSF money for that

Finally, computer science’s notion of acceptable methodology has changed significantly in our lifetimes. It was a long hard fight to get HCI accepted as a valid field of computer science. And before that, it was a long hard fight to get Computer Science accepted as science at all.⁴ Many of us still have colleagues who don’t view AI as science.

This isn’t to say that science is phantasmal or that method is unimportant. Method matters. But we need a more sophisticated debate on method. AI-based narrative involves a range of different projects from modeling cross-cultural aspects that abut cognitive psychology (Ware & Young, 2012) to the production of technologically sophisticated artworks (Mateas, Domike, & Vanouse, 1999; Mateas & Stern, 2005). Such different projects require fundamentally different methods with

³ It’s interesting to note how much more often the term is used in the negative than the positive. We frequently hear papers declared **not** to be science, but I’ve yet to hear someone jump up in a talk and declare “that’s science!”

⁴ Imagine trying to get your work respected if there was no computer science department at your university and you were instead inside the math department.

fundamentally different notions of rigor. This is best served by an open and energetic debate about method. The tongue-in-cheek title of this paper isn't actually an objection to science itself but rather to invocations of science that serve to shut down that debate.

From data to cases: rigor in AI-based art

Not everyone in the INT community is developing technologies for art and entertainment systems. But for those who are, it should be perfectly appropriate to make an artwork that uses the technology. That artwork should be developed and evaluated using the best methods available, including the methods used by practicing artists, such as studio critique.

Statistical methods (user studies, analytics etc.) can certainly be useful. Analytics-driven game companies such as Zynga have been very successful hill-climbing existing designs through relentless A/B testing (Pincus & Gordon, 2009). But they've also demonstrated that there are severe limitations to what you can do with those kinds of statistical measures. The dimensionality of the design space is too high to have any hope of sampling it in a systematic way. So the most you can hope to do with point-measurements in that kind of high dimensional space is hill climbing.

But anyone who has taken a course in painting, writing, film making, musical composition, or any other art has probably noticed that they weren't trained to make art as if they were focus-grouping a new product. That's not because of an inherent bias against scientific methods, but simply because the design space is so impossibly large you couldn't possibly gather enough data to be able to know what the optimal haiku was, much less the optimal arc for a multi-decade soap opera. Plus, my optimal multi-decade soap opera would be different from your optimal multi-decade soap opera because different people value narratives differently.

If you can't gather data across all possible designs and users, then the alternative is to look as deeply as possible at designs you do examine; to glean as much as possible from a specific encounter with a specific piece. So art training and theory are traditionally focused on examining specific pieces, specific cases. And as such, it is in some ways closer to the practice found in law and business than in engineering.⁵

Artists are trained to develop an awareness of how a given piece affects them; to watch themselves watching it; to develop an awareness of how a given attribute of a given work affects their perception and interpretation of it *in that*

case. Although the goal is in some sense to develop a sufficient understanding to be able to invert the mapping, and predict how a hypothetical design might affect a viewer's experience, that understanding is nonetheless always partial and contingent, so the actual creation process is iterative, with each iteration forming a case to be examined and used in the formation of the next iteration.

This presents a serious challenge for AI work in support of the arts because computer science is traditionally, and understandably, resistant to work on cases. One cannot generally get a Ph.D. for writing a program; one gets a Ph.D. for developing a general theory that will help others write some class of programs. But I see no alternative but to relax this constraint for AI work that have specifically aesthetic goals. We have to be open to the case because the case is all we have.

Conclusion

My concern is that we're at risk of letting our methods drive our inquiries rather than letting our inquiries drive our methods. I have on several occasions been told by students at conferences that they would like to make a game or artwork that demonstrates the ideas from their thesis but that "that's not science."

That makes me want to shoot myself.

If we limit ourselves to using only analytics and psychology-style human subjects experiments, we are at best crippling ourselves and at worst deceiving ourselves with a false sense of rigor. In the worst case, we could end up creating (unironically) a kind of interactive narrative version of Komar and Melamid's *People's Choice* series (Wypijewski, 1997) in which they commissioned a market research company to study the characteristics most- and least-wanted in paintings by different demographic groups, then painted the results. Their *United States: Most Wanted Painting* (1994) is a realist landscape incorporating trees, water, mountains, George Washington, a happy family, and a pair of deer.

I recognize there is *Realpolitik* to science. We must all make ourselves intelligible and respectable to our departments, Deans, hiring committees, and funders; and that may require us to use methods and metrics we don't always believe in. But that doesn't mean we have to limit ourselves to *only* using those methods.

As Feyerabend (2010) argued, what matters is forward progress; any method that helps make progress is worthwhile. And just as early HCI researchers worked to establish their methods as valid computer science, we should work to establish the ones we believe in too.

In the meantime, I look forward to the day I stop hearing the phrase "I'd like to do X, but that's not science."

⁵ This is of course an oversimplification. The use of the cases varies greatly between fields. See Berlant (2007) for a discussion of the use of cases in different disciplines.

References

- Abbott, H. P. (2008). *The Cambridge Introduction to Narrative* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Barthes, R. (1966). Introduction à l'analyse structurale des récits. *Communications*, 8(8), 1–27.
- Berlant, L. (2007). On the case. *Critical Inquiry*, 33(Summer), 663–672. Retrieved from <http://www.jstor.org/stable/10.1086/521564>
- Eagleton, T. (2008). *Literary Theory: An Introduction* (3rd ed.). Minneapolis, MN: University of Minnesota Press.
- Feyerabend, P. (2010). *Against Method* (4th ed.). New York, NY: Verso.
- Horswill, I., Riedl, M., & Young, R. M. (2011). Directions From Here: Questions for Intelligent Narrative Technologies (panel discussion, October 11). In *Fourth Workshop on Intelligent Narrative Technologies*. Stanford, CA.
- Komar, V., & Melamid, A. (1994). United States: Most Wanted Painting. Retrieved from <http://awp.diaart.org/km/usa/most.html>
- Lakoff, G. (1990). *Women, Fire, and Dangerous Things*. Chicago, IL: University Of Chicago Press.
- Mateas, M., Domike, S., & Vanouse, P. (1999). Terminal Time: An Ideologically-biased History Machine. In *Proceedings of the AISB '99 Symposium on Creative Language: Humor and Stories* (pp. 69–75).
- Mateas, M., & Stern, A. (2005). Façade.
- Pincus, M., & Gordon, B. (2009, October 28). A Serious Take on Internet Game Play. *Stanford University Entrepreneurial Thought Leader Lecture Series*. Stanford, CA, USA.
- Schank, R., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum Associates.
- Todorov, T. (1969). Grammaire du “Décameron.” In *Approaches to Semiotics 3*. the Hague: Mouton.
- Ware, S. G., & Young, R. M. (2012). Validating a Plan-Based Model of Narrative Conflict. In *International Conference on Foundations of Digital Games*. Raleigh, NC.
- Wittgenstein, L. (1958). *Preliminary Studies for the “Philosophical Investigations”, Generally known as The Blue and Brown Books*. Oxford, UK: Blackwell Publishers Ltd.
- Wypijewski, J. (Ed.). (1997). *Painting by Numbers: Komar and Melamid's Scientific Guide to Art*. New York, NY: Farrar Straus Giroux.