

### 3.3 Interdisciplinary Research Methods

*Ian Horswill (Northwestern University – Evanston, US)*

**License** © Creative Commons BY 3.0 Unported license

© Ian Horswill

**Joint work of** Horswill, Ian; Mateas, Michael; McCoy, Josh; Paiva, Ana; Rudolph, Günther; Smith, Gillian; Young, Michael; Zhu, Jichen

Game AI research is inherently interdisciplinary, involving not only computer science, but also design, psychology, sociology, graphic design, theater, filmmaking, creative writing, and a host of other areas. A system can be successful from a technical computer science standpoint, while still being an abject failure as a piece of art or entertainment. There is thus an open question as to the appropriate ways to evaluate Game AI research contributions.

While the traditional evaluation methodologies of computer science, including HCI methodologies such as controlled user studies, are undeniably useful, we will likely need to look to other disciplines, particularly in the arts and humanities, for methods for evaluating the aesthetic, thematic, or ideological dimensions of a work.

Evaluation of this kind of interdisciplinary work is complicated by the differing notions of problem adopted by the various disciplines. Engineers, including computer scientists, typically work on well-posed problems – problems with a clear formulation and natural performance metrics. Designers on the other hand, often work on so-called “wicked problems,” ones without definitive formulations or metrics, and little ability to generalize solutions across problems. Artists may not even conceptualize their work as solving a problem, but merely as exploring an interesting region of design space to see what they encounter within it.

Another complication lies in the different intellectual and social purposes of evaluation. Although evaluation sections of papers are often written as if answering the question of how well a system worked, they are often read by reviewers and other gatekeepers as answering the question of how well the researchers worked, and thus, whether the work should be published. And yet, they might be most usefully written and read as trying to articulate what insights can be derived from the system for the guidance of future system builders.

We can find resources for evaluation from the component disciplines. The use of performance metrics and correctness proofs from engineering; quantitative methods such as survey instruments as well as qualitative methods such as ethnography from the social sciences; artist talks, critique sessions, and juried shows from the fine arts; postmortems and commentary tracks from the game industry; close reading and critical theory from the humanities; symbiotic interaction between critical reflection and algorithmic development from critical technical practice; and the use of the case study, common in areas such as medicine, law, and business.

However, we don’t currently have good evaluation methods for characterizing many of the issues that come about in game AI, and acknowledge the need to create new evaluation methods. How do we evaluate architectures as opposed to systems? How do we evaluate generation systems in PCG? Or the authorial effort involved in using a particular system?

Understanding these issues is important to the future development of the field. We look forward to a rich dialog on evaluation, both within the field of game AI and in conversation with sibling disciplines from which it can learn.