ACM INTERACTIONS, NOV/DEC 2022. DOI: 10.1145/3564038

Feature

What Simulation Can Do for HCI Research

Roderick Murray-Smith*, Antti Oulasvirta, Andrew Howes, Jörg Müller, Aleksi Ikkala, Miroslav Bachinski, Arthur Fleig, Florian Fischer, Markus Klar

Correspondence
*School of Computing Science, 18 Lilybank Gardens, University of Glasgow, Glasgow G12 8RZ.
Email: Roderick.Murray-Smith@glasgow.ac.uk

Insights:

- Simulation-based methods are developing rapidly, changing theory and practice in applied fields.
- Simulations help create and validate new HCI theory, making design and engineering more predictable, and improving safety and accessibility.
- Emulation of user behaviour with generative models tests our understanding of an interactive system.
- Simulation-based intelligence can be created by directly embedding models in intelligent interactive systems.
- Model-based evaluation provides insights into usability before user-testing, saving money, time, and discomfort.

1 INTRODUCTION

The early days of every engineering subject involve examples of expensive failure when the skilled artisans of the day succeeded in gradual progress, but these successes were punctuated by the disasters which occurred when they made too large an innovation step. From the collapse of cathedrals in France, the capsizing of the 17th-century Swedish warship Vasa, and more recent failures including air accidents attributable to modern cockpit designs, we see the potential high cost of 'in-the-wild' prototyping approaches, especially in modern environments involving rapidly changing demands, or where the complexity and expense of prototyping increases significantly.

While long-established in fields such as civil engineering, naval architecture and aeronautics, use of modelling and simulation to test designs long before physical prototypes are created, simulation-based design methods are developing rapidly, expanding to new fields such as pharmaceutics, epidemiology and medicine, coupling formal models of scientific theory with large-scale data acquisition to calibrate the models. In this paper we argue that simulation can aid in such large technological strides forward, but also can support design 'in the small', especially with often underrepresented user groups or contexts.

Human-computer interaction research and practice has been slow to adopt simulation, in part because many have argued that traditional human-based usability testing is quicker and more valuable than offline simulation. But the ability to build a generative model that matches user behavior is a strong test of whether we understand an interactive system. Furthermore, simulations can support the creation and validation of new theories, and make design and engineering more predictable and robust processes. Simulation can be directly embedded in intelligent interactive systems, and has the potential to improve system safety, and accessibility. Model-based evaluation can provide insights into usability *before* testing with end users, and we argue that in the future, in many cases, this cost in money, time or human discomfort of doing extensive parameter optimisation via experiments with human participants will make it impractical or unethical to avoid the use of simulation.

2 WHAT IS SIMULATION IN HCI?

A model of a system, artefact or environment is a simplified representation that captures its essential characteristics for a specific purpose. A simulation is the operation of the model, where the intention is to draw conclusions, qualitative or quantitative, about the behaviour or properties of a real-world process or system over time. Simulation is an indispensable tool for scientific research wanting to understand the behaviour of complex systems, including hypothetical, extreme or dangerous conditions, or situations where it is too slow or expensive to use the real-world process itself (Lavin et al., 2021). It allows an unambiguous implementation of the current scientific theory, and predictions can be validated with observed real-world data. Any mismatches prompt researchers for the requirements for the next steps in theory development, or data acquisition.

Typically, performing a simulation of a system means using the execution of a computer program to approximate the behaviour of a mathematical model. In a broader sense, the simulation is a method for studying systems and their behaviour, which includes choosing a model; finding a way of implementing that model in a form that can be executed on a computer; running the model to compute the outputs; inverse applications of models to infer hidden states or parameters; validating the model; visualising, analysing, and interpreting the resultant data to find explanations.

In HCI, pioneering work in simulations was driven by (Card et al., 1983), whose Model Human Processor (MHP) divided the user aspect into cognitive, motor-behavioural and perceptual components. They introduced GOMS (Goals, Operators, Methods to achieve the goals, Selection of competing methods) models to predict task times based on separating tasks into elementary events and summing the expected time to complete the user task.

Criticisms of the use of simulation in HCI have included their cost and complexity to develop, and their inability to adequately represent the cognitive and perceptual complexity of the human in the HCI loop, especially given the sensitivity of behaviour to details of context. Other critiques have included the perceived failure of models to capture the physical and social context of interaction. In contrast, we argue that investing in simulation models will actually save expenses and time, by streamlining the development process.

3 SIMULATION FOR THEORY-FORMATION

Simulation helps push theory forward by virtue of the fact that it demands abstraction. Simulation does not require the veridical replication of the movement of every molecule, rather it must be based on appropriate abstractions that form the ‘units’ of simulation. The choices of these abstractions go hand-in-hand with theoretical development. In cognitive science, for example, artificial neural networks are often considered as simple simulations of aspects of biological neural networks. Other phenomena may require different theoretical commitments and radically different kinds of simulator – e.g. spiking neural networks.

In HCI and in cognitive architectures a commitment was made to Fitts' Law because it was felt that stochastic submovements – processes that give rise to the law – were unimportant for predicting pointing performance. Fitts' law was used as a base-level abstraction in MHP, for example. However, this assumption has proved too restrictive for explaining key phenomena concerning adaptation, and alternative simulation environments are now available . For example, full-body biomechanical simulations using Reinforcement Learning are used to predict not only movement performance but motion trajectories and even fatigue during pointing (Cheema et al., 2020).

This coupling of the appropriate abstraction and precision in defining a simulator is also critical in grounding experimental work. (Thimbleby, 1990) observed, *"We can do as many 'experiments' as we like on complex systems, evaluating systems with vast numbers of people, doing sophisticated statistical tests, and so on, all to no avail unless we know what we are doing, and how the results of the experiment bear on future work. [...] I search the literature for theories that I can apply in my case [...]; instead I find reports of experiments – sometimes related to my particular problem – but without some underlying theories, how can I know how safely I can generalise those results to apply in my design, with my users?"*.

However, one lesson that HCI has learned over the past 50 years is that *the details matter*. The context matters, the user matters, and both can change behaviour significantly. Understanding and describing the variability of human behaviour, and sensitivity to context is a significant challenge, but one that can be supported by simulations.

4 SIMULATION FOR DESIGN AND ENGINEERING

Can we advance user-centred design to a more rigorous, safer, and predictable process via a simulation-based approach to interaction design? Simulations can be used to predict task performance, such as time taken to finish the task and how often tasks can be successfully accomplished, e.g. keystroke-level models in (Card et al., 1983). They can be biomechanical models, predicting movement, its physical ergonomics, physiological and health effects. They can have components predicting perceptual performance in different contexts, and they could include cognitive elements. Offline simulations can improve robustness through more thorough exploration of the design space.

Many interactions have already been modelled with biomechanical models: a button press, touch on a screen, mid-air gestures, up to full-body movements (Cheema et al., 2020). These interactions can be evaluated by comparison with human data, e.g. from motion capture. Simulations of such movements are computationally demanding, but today various suites of dedicated biomechanical simulation software can be found to efficiently carry out these computations, e.g. OpenSim,[1] AnyBody, LifeModeler, and SantosHuman, as well as powerful physics engines with biomechanical modelling capabilities (e.g. MuJoCo[2], Bullet[3]). Biomechanical simulations can include *inverse simulations* which enable inference of hidden states or parameters from experimentally observed movement data, and forward simulations which predict complete movement behaviour.

It will often be impossible to avoid simulations because of the cost in both money and time of doing extensive parameter optimisation via experiments with human participants. For example, in (Williamson et al., 2010) a bearing-based pedestrian navigation system had parameters such as the size of the angular window needed for feedback when pointing at the target. A simulator modelling pedestrians as rational agents optimised the ideal window size for efficient navigation under different assumptions of sensor uncertainty – an extremely time-consuming task for multiple large groups of human participants. (Kristensson & Müllners, 2021) use modelling to replace extensive experimentation, optimising text entry system parameters.

[1] `https://simtk.org/projects/opensim/`
[2] `https://mujoco.org/`
[3] `http://bulletphysics.org/`

5 SIMULATION FOR INTELLIGENT INTERACTIVE SYSTEMS

Previous sections described the role a simulation can have *offline*, during the design stage. However, simulations can also run *online* in real-time, or faster-than-real time, with their initial states based on current conditions. This can allow a system to act as a 'digital twin' monitoring activity, and inferring hidden states (such as possible user intentions or goals), or it can predict possible outcomes and use this information to adapt the interface, make decisions, or change the feedback to the user.

The ability to perform faster-than-real-time simulation allows predictive interfaces to offer auto-complete options, or to jump to likely targets, or permit more sloppy actions from the user (Weir et al., 2014). In more safety-critical applications it could be used to provide the user with warnings about the consequences of their current behaviour, if there is a possibility it will lead to dangerous states.

6 SIMULATION FOR SAFETY, ACCESSIBILITY AND ETHICAL REASONS

Traditional user testing has its place, but as systems become more complicated, and users more diverse, we hit complexity, robustness, and ethical challenges in usability testing. (Quek, 2013) observes, *"In contrast, it can be more difficult to access people with disabilities and a user study can take longer and is more effortful for the participant. Inclusive design aims to make it possible for mainstream applications to be used by people of all abilities. To this end, good models and guidelines must be made available to designers and developers."*. In the near-future it may be deemed unethical in some domains, such as those with vulnerable users, to propose an experiment with human users before every effort has been made to reduce the uncertainty about the outcome with other means. A key element will be a rigorous simulation of the experiment.

Simulation can be necessary for a number of reasons: It may be too risky to test a system without initial simulation of proposed experimental conditions. The risks can be physical, emotional or ethical. For niche user groups, the availability of the users locally may be very limited, and it may be difficult to persuade participants to take part in multiple trials (or not in line with experiment protocol). Use of simulation forces designers to be explicit about the elements included in the model, making the design process more auditable for stakeholders, e.g. to identify underrepresented users.

When designing for inclusion, an empathic modelling approach focuses on simulating a disability to allow designers to understand a system from the user's point of view, and better appreciate the problems the system should tackle, allowing a narrowing down of options, and limiting fatigue and frustration for the participants. (Quek, 2013) provides a review of the literature of simulation as part of the evaluation process for vulnerable groups, and develops a specific example of the use of simulations of Brain Computer Interfaces to explore design options and allow able-bodied users to test the interaction before disabled users were asked to test the system. This approach also allowed the creation of multi-participant software such as games, where people of different input abilities and input mechanisms could be placed on an equal footing, by using the simulations to create a common denominator among all users.

For safety reasons, simulation has long played a key role in aviation in training pilots, testing new flight procedures, or aircraft design changes. Similarly, as autonomous driving and associated interfaces grow in importance, we anticipate an expanded need for simulation in UI-design for automobiles. In general, deploying untested systems to millions of users is highly risky in terms of reputation, customer retention and longer-term financial consequences. E.g., in recommender systems, a gap has opened between research and practice, due to the vulnerability of the traditional approach of testing on historical logs of user interactions. Recommender systems which perform well on historical data often rapidly go wrong when they engage with real users. While A/B testing with population subsamples can reduce risks, closed-loop simulations based on user models can be used to pre-test the system before user-testing.

7 WHY NOW?

Why do we believe that the time is ripe to reconsider what models can do for HCI? (Lavin et al., 2021) present a far-reaching topical review of the role of simulation in science and AI (but few HCI examples), highlighting that in the past, the complexity of simulation was constrained by hardware limitations, the lack of information, the difficulty of dealing with uncertainty, and practical challenges in integrating multiple different simulation models, reducing the utility of the simulation approach for practical decision-making. However, recent developments in probabilistic, differentiable programming, high-performance computing, causal modelling, and the rapidly improving ability of machine learning to emulate complex aspects of human perception and behaviour mean that simulation has an increased potential to be efficiently, and usefully applied in new domains such as HCI (Jokinen et al., 2017).

New sensor technologies and new interaction styles, such as augmented reality, take us from our comfort zone with well-understood mechanisms, because of issues such as the sensor fusion, high-dimensional and uncertain sensors, and the application of machine learning technology for segmentation and labelling content. The increased complexity of system design will demand more use of simulation in its development and optimisation.

A common challenge, relevant to HCI, is linking simulation-based models with empirical data. Numerical simulators typically have parameters whose values are not known *a priori*, and have to be inferred by data. Classical statistical approaches are not always easy to apply to models defined by numerical simulators, and in some cases simulations may be in legacy code, or only available as black-boxes. A key recent development is *Simulation-Based Inference (SBI)*, also known as *likelihood-free inference*, which enables researchers to algorithmically identify parameters of simulation-based models that are compatible with observed data and prior assumptions.

8 OUTLOOK

We have highlighted that developments in hardware and software can enable us to create increasingly complex models of human interactions with technology, and calibrate them to increasingly rich and available data. An example is the application of methods from artificial intelligence and machine learning, which are having a wide-reaching impact on many areas of day-to-day life and in science. We believe that the natural outcome of the application of these technologies will be in simulation systems that can have a combination of first-principles, white-box modelling and flexible, data-driven black-box models. It is important that stakeholders with an understanding of specific user groups can be involved in the specification and evaluation of these models.

Simulations can be used *offline*, during the design process, to reduce stress on vulnerable user groups, increase the rigour and reproducibility of testing, ensure diversity in testing processes, and improve safety. They can speed up design and development, and reduce project development time uncertainty.

Use of simulations *online* allows us to include a predictive element in computational design of interfaces, and explore multiple scenarios compatible with observed data. Such online simulation of human behaviour is likely to be a core requirement of any future ‘intelligent’ interactive systems.

In addition to these practice-focused improvements, simulation can help the scientific process in HCI research. The need for formal rigour in creation of a simulation model, and controlling and documenting the provenance of data used to calibrate it, makes clear the importance of many of the often poorly described aspects of context in HCI experiments. A simulation package is also easily shared with other researchers, improving reproducibility.

Aspects of models which at any given moment in time are poorly justified theoretically, are a poor fit to experimental data, or which are highly sensitive to context can be viewed as prompts to the research community about where they need better theories, more complex models, or more data. This can create a shared awareness of the open problems and challenges, and can help document progress.

Lord Kelvin said, *"When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind."* We argue that if we cannot build a generative simulation model which can, to some useful degree, replicate aspects of the complexity and variability of human behaviour in a given interaction context, then our knowledge of the expected interaction and its consequences, is still of a 'meagre and unsatisfactory' kind. However, if we have a concrete simulation model with known weaknesses that need improving, then we at least know where to begin to develop our theory and acquire more data, in order to rectify that 'unsatisfactory state of affairs'.

Card, S. K., Moran, T. P., and Newell, A. *The psychology of human-computer interaction*. Lawrence Erlbaum, 1983.

Cheema, N., Frey-Law, L. A., Naderi, K., Lehtinen, J., Slusallek, P., and Hämäläinen, P. *Predicting Mid-Air Interaction Movements and Fatigue Using Deep Reinforcement Learning*, pp. 1–13. Association for Computing Machinery, New York, NY, USA, 2020. ISBN 9781450367080.

Jokinen, J. P. P., Sarcar, S., Oulasvirta, A., Silpasuwanchai, C., Wang, Z., and Ren, X. Modelling learning of new keyboard layouts. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 4203–4215, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450346559.

Kristensson, P. O. and Müllners, T. Design and analysis of intelligent text entry systems with Function Structure Models and Envelope Analysis. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2021.

Lavin, A., Zenil, H., Paige, B., Krakauer, D., Gottschlich, J., Mattson, T., Anandkumar, A., Choudry, S., Rocki, K., Baydin, A. G., et al. Simulation intelligence: Towards a new generation of scientific methods. *arXiv preprint arXiv:2112.03235*, 2021.

Quek, M. *The role of simulation in developing and designing applications for 2-class motor imagery brain-computer interfaces*. PhD thesis, University of Glasgow, 2013.

Thimbleby, H. *User Interface Design*. ACM Press, NY., 1990.

Weir, D., Pohl, H., Rogers, S., Vertanen, K., and Kristensson, P. O. Uncertain text entry on mobile devices. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 2307–2316, 2014.

Williamson, J., Robinson, S., Stewart, C., Murray-Smith, R., Jones, M., and Brewster, S. Social gravity: A virtual elastic tether for casual, privacy-preserving pedestrian rendezvous. In *Conference on Human Factors in Computing Systems - Proceedings*, volume 3, 2010. ISBN 9781605589299. doi: 10.1145/1753326.1753548.