Does Simulation Fidelity Affect Training? A Lesson from a Brief Review of Literature

Pooyan Doozandeh and Frank E. Ritter
pooyan.doozandeh@gmail.com frank.ritter@psu.edu
College of Information Sciences and Technology, Pennsylvania State University
University Park, PA, 16801, USA

Abstract

The surface realism of training simulators, or fidelity, has been widely assumed to affect training. As such, fidelity has had a major role in guiding the research and design of training simulators. However, there is a growing body of evidence that questions the role of fidelity in the effectiveness of training. Here, we briefly present the challenge of research on fidelity. The review of literature indicates that fidelity—to the extent that it refers to the surface realism—can no longer guide the research and design. Based on this lesson from the literature, we call for a new theoretical framework that would replace fidelity. The new framework should provide specific technical recommendations that would guide the research and design of training systems.

Keywords: Simulation Fidelity, Surface Realism, Transfer of Training

1 Introduction

People need training before doing many tasks. To reduce the time and cost of training, as well as removing the associated risks of practice in certain domains, trainees practice tasks with simulators that are simplified replications of target tasks. Nowadays, computerized training simulators are used in a wide variety of domains for training pilots, soldiers, firefighters, and physicians [1-3]. The goal of a training simulator is to train a skill that is required for the target task. The effectiveness of a simulator thus depends on the extent to which acquired skills through practicing the simulated task can be transferred to the target task. So, transfer indicates the effectiveness of training.

The question is how to build effective training simulators? Specifically, what factor(s) determine the transfer of training in a simulator? There are several factors that researchers and designers often manipulate to assess their effect on transfer. One of the factors that has repeatedly been the subject of research and investigations is the level of the faithfulness of simulation to the target task, or simulation fidelity [4]. As Table 1 shows, fidelity has been used to refer to various aspects of the faithfulness of simulation, and for this reason, researchers specified different types of fidelity.

Notwithstanding this type-specific use of fidelity, when fidelity is used without the specification of the type and as a holistic and unified concept, it refers to the degree of faithfulness of simulators to the physical and perceptual aspects of target tasks. In this way, Hays and Singer [11] framed the *fidelity question* as follows:

> how similar to the actual task situation must a training situation be to provide effective training? [11]

This question has a long history in discussions on training systems [6, 12-13]. Investigations on fidelity often address the extent to which the details of the simulated environment (i.e., control panel, visual scene, and other observable aspects of a training simulator) should resemble those of the target task environment.

Fidelity, as a construct of research and design, has had a major role in guiding the design and improvement of training simulators over the last few decades. In particular, the progress in computational systems and simulations after the early 2000's provided the opportunity to create realistic simulations of target task environments in simulators. However, there has been a growing body of evidence that questioned the effect of fidelity on the effectiveness of training. The question is, has the increased realism of simulation resulted in higher levels of transfer in past studies?

Table 1. Examples of type-specific uses of fidelity.

Type of fidelity	Definition
Physical	"the attempt to represent accurately the appearance and "feel" of the actual equipment" [5-6].
Task	"refers to the correspondence between tasks performed on the actual equipment and tasks performed on the training simulator" [7].
Behavioral	"the [description of] operator's behavior by system identification methods … If two systems are behaviorally equivalent, we might assume that training in one would transfer positively to the other" [8].
Environmental	"the degree to which the simulator duplicates the sensory stimulation (excluding control feel) which is received from the task situation" [4, 9].
Equipment	"the degree to which the simulator duplicates the appearance and "feel" of the operational equipment" [4].
Psychological	"the degree to which the simulator is perceived by the trainee as being a duplicate of the operational equipment and the task situation" [9, 6].
Functional	"the attempt to represent faithfully the stimulus and response options provided by all or portions of a piece of equipment" [10].

2 Review of Literature

We explored the literature to find theoretical and empirical works that, directly or indirectly, addressed the relation between fidelity and the effectiveness of training. The criterion of selection of papers in this review was the theoretical or empirical relevance of prior research on the subject.

2.1 Traditional Theory: Fidelity-Transfer Correlation

After World War II, mainly because of the technological advances and investments in military, there was a growing need for training personnel and recruits for various situations, including operating advanced machineries, flying aircraft, and special physical and situational skills (for a review, see Blaiwes & Regan [14]). In creating training simulators, one of the design factors that became the focus of attention was the degree of simulation, or fidelity [4]. It was hypothesized that fidelity of simulation could affect the transfer of training from simulation to the target task environment.

In a classic and highly influential work in the field, Miller [6] presented a detailed analysis on the psychological considerations in designing training systems. Among many of the topics covered in Miller's article, one theory seems to have had a long-standing impact on the community of researchers and designers of training simulators. As outlined in Figure 1, Miller formalized the relationship between the degree of simulation and the transfer of training. If we translate Miller's words to today's terminology of simulation, the horizontal axis in Figure 1 indicates the fidelity of the training simulator, and the vertical axis shows the effectiveness of training as indicated by the degree of transfer.

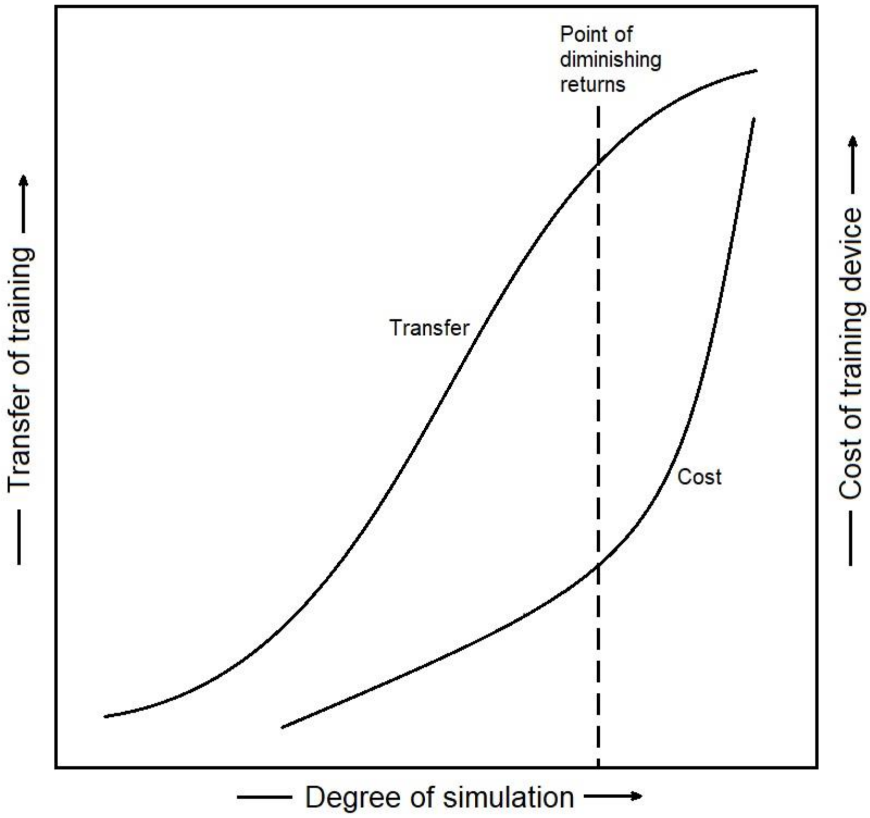

Fig. 1. Miller's [6] depiction of the relationship between the degree of simulation (i.e., level of fidelity, on the horizontal axis) and the transfer of training (vertical axis).

The hypothesis was that more effective simulators should be those with higher fidelity. But because high-fidelity simulators were costly in that time, the problem was in finding the most efficient level of fidelity in training systems. The "point of diminishing returns" in Figure 1 is the level of fidelity that was argued to be the optimal level for training simulators (for an early similar discussion in diminishing returns of training for pilots, see also Gagne [15]).

In a similar attempt more than thirty years later, Alessi [12] refined Miller's curve by distinguishing between trainees based on their expertise. Specifically, Alessi argued for the difference in the effect of fidelity between novice and expert trainees. As shown in Figure 2, high-fidelity simulators are effective only for expert trainees; students and novices do not benefit as much from high-fidelity simulation. However, Alessi did not discuss the requirements to build simulators for novices. This is crucial because many of the training simulators are used by trainees who do not have the expertise and experience on the task.

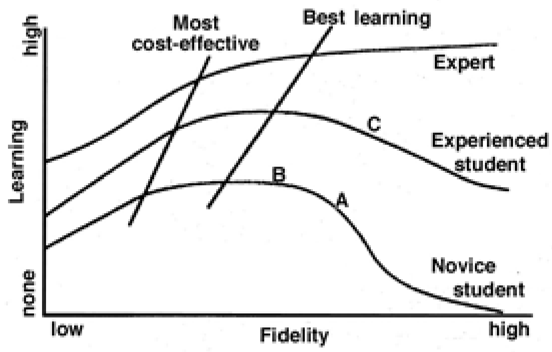

Fig. 2. Alessi's [12] hypothesis for the relationship between the level of fidelity and training effectiveness for trainees with three levels of prior expertise.

As a result of Miller's and Alessi's works, as well as other research with similar theories [16], fidelity—as the degree of realism of simulation—became the central design factor that engineers and researchers could manipulate in designing training systems. In this tradition, subsequent studies [11, 17-18] worked on Miller's hypothesis by considering more contextual factors in determining the appropriate level of fidelity for training systems (e.g., trainees' level of expertise, training stage, the task). The question of developers and deployers of training simulators then became: "*how similar to the actual task situation must a training situation be to provide effective training?*"—what Hays and Singer [11] called the *fidelity question*. Miller's theoretical framework, together with subsequent investigations, initiated a line of research that aimed at determining the most optimal and cost-effective level of realism for simulators.

For instance, Parrish, McKissick, and Ashworth [19] provided empirical evidence that showed high-fidelity flight training simulators resulted in improvement in pilots' performances. In the same domain, Noble [20] supported Alessi's hypothesis by arguing for the existence of an optimal point of fidelity level in the training of non-expert pilots. In the domain of nurse education, Cant and Cooper [13] found the effectiveness of medium- and high-fidelity simulations in training nurses, and their finding has widely been used and implemented in medical training simulators [21, 22]. Similarly, in maintenance training systems, Allen, Hays, and Buffardi [5] tried to show that fidelity was correlated with training outcome.

In summary, by accepting Miller's hypothesis, many of the empirical investigations over the last few decades tried to determine the optimum level of fidelity for training in various domains. Miller's hypothesis and subsequent theories constituted what we would refer to as the *traditional theory*. This theory is based on two main premises:

> *There exists a relation between the fidelity of simulation and transfer of training; and*
> *There exists a positive correlation between fidelity and transfer.*

It seems that accepting the correlation between the level of fidelity and transfer, or *fidelity-transfer correlation*, became a standard in many of the subsequent empirical research. In fact, addressing the fidelity question assumes the fidelity-transfer correlation. These assumptions of the traditional theory, however, were challenged by subsequent researchers.

2.2. Challenging Findings

In one of the early critical articles in the field, Fink and Shriver [23] reviewed the maintenance training systems that were used during the 1950's and 1960's, and argued that, at least in certain maintenance tasks, using low-fidelity simulations are helpful and sometimes necessary. In a more recent review of maintenance training systems, Swezey, Perez, and Allen [24] found that using high-fidelity training of the electromechanical troubleshooting tasks (i.e., animated training materials) did not improve the training outcome when compared with low-fidelity training systems for the same tasks.

Challenging the faith in high-fidelity simulators was not limited to the field of maintenance and troubleshooting. Havinghurst, Fields, and Fields [3] compared high- and low-fidelity simulations for training firefighters and observed insignificant differences in the performances between the two modes of training. In health care and medical training, although the belief in the traditional theory still strongly exists, Beaubien and Baker [25] argued that researchers assumed the effectiveness of high-fidelity simulation, and this masked the possible benefit of using low-fidelity simulators (see also Hamstra, Brydges, Hatala, Zendejas, & Cook [2]).

The faith that the medical training community continues to hold in the effectiveness of high-fidelity simulations [21] was dubbed "naïve realism" by Smallman and St. John [26]. With respect to visual interfaces, Smallman and St. John argued that using "spartan" displays and graphical images can lead to more effective representations in computer simulations. Therefore, despite the intuitive but naïve appeal of realistic displays, Smallman and St. John espoused the idea that low-fidelity graphical displays can be advantageous in various areas of simulation. This is mainly because low-fidelity graphical displays could focus on the aspects of a scene or environment that are more relevant for specific tasks. The effectiveness of using low-fidelity graphical representations in training was also shown in a recent study by Cöltekin et al. [27] in which it was shown that using low-fidelity abstract geographical maps of an environment improved map-based route learning compared to when subjects used realistic high-fidelity maps of the same environment.

Finally, in the domain of flight-training simulation, as one of the most important areas that use training simulators, Dahlstrom, Dekker, van Winsen, and Nyce, [28] showed how experienced pilots favored using mid-fidelity flight simulators that lacked the realism of high-fidelity simulation. Their finding questioned the efficiency of realistic flight simulators as they are widely used for training. This is of particular importance because high-fidelity simulators in fields such as aviation are costly, and so, designers should always look for ways to reduce the cost of training systems.

The studies mentioned above are examples among a growing body of evidence that are questioning the traditional theory by challenging the fidelity-transfer correlation. This challenge was brought mainly by showing how low-fidelity training simulators could result in comparable training outcomes—and in some cases, better outcomes—than high-fidelity simulators. In the next section, we briefly review the possible implications of these findings that questioned the validity of the traditional theory.

3 Conclusion: Questioning the Fidelity Question

The perspective described above divides the history of fidelity in training simulation in two parts: the dominance of the traditional theory, and the emergence of challenging findings. The traditional theory was formed as a response to an increasing need for a practical guideline in designing training simulators, and it proposed fidelity as the primary design factor. By believing the fidelity-transfer correlation, the followers of the traditional theory argued that to increase transfer, researchers and designers had to manipulate surface characteristics of simulation; the more realistic the simulation looks and feels, the higher the transfer of training would be. The outcome of this perspective was the research on answering the fidelity question: how much fidelity is needed to produce the best training outcome. This theory subsequently initiated a line of research that engaged researchers and designers, and has had practical implications on the design of training systems over the last few decades.

However, subsequent empirical findings questioned the validity of fidelity-transfer correlation. These empirical studies countered the intuitive appeal of using realistic simulations in training, and argued against the necessity of using high-fidelity interfaces for better training outcomes. The question that mainly guided the challenging findings was whether low-fidelity simulators could have similar training outcomes as their high-fidelity counterparts; they not only showed this possibility [24], but also presented results on how some low-fidelity systems could be more effective than high-fidelity ones [27]. This possibility is especially important in fields that utilize costly training simulators (e.g., flight training). More importantly, these findings call for revising the belief and usage of the traditional theory. If the fidelity of simulation cannot accurately determine the training outcome, investigations on improving simulators by manipulating their fidelity will no longer be a valid project.

From one perspective, recent studies pose a challenge to the argument that fidelity can determine training outcome. However, the literature shows that whether fidelity is correlated with transfer depends on the specific field and task, trainees' level of expertise, and various other contextual factors [12]. Therefore, the important lesson from the literature is not whether high-fidelity simulators are effective and useful; rather, it is the growing disbelief on fidelity that can be seen among researchers. Manipulating the surface realism with the goal of designing or improving training simulators is no longer popular among researchers and funding organizations. Because of this diminishing popularity and to reduce further confusions, some authors in recent years proposed to stop using the concept of fidelity in all discussions on training systems [2].

This growing disbelief in fidelity can have various implications. As a result of this situation, the fidelity question (i.e., what level of fidelity is needed in a training simulator?) would no longer be a legitimate question to ask because it cannot produce novel findings. In a broader level, because fidelity has guided the research and design of training system for decades, removing fidelity would cause a void. This is because recent challenging findings questioned the fidelity, but did not offer an alternative to replace it in research and design. This void of not having a reliable construct for research and design is also the reason that after about five decades, military organizations are still funding researchers to produce practical guidelines for the design of training systems. Research projects funded by the armed forces during the 1950s and 1960s [6, 29-30] looked for similar design guidelines as their recent equivalents [31-33]. In short, the design community currently lacks a resource that provides guidelines of how to design and improve training simulators, and current attempts are mostly based on intuitive hypotheses and unverified practices. The question that still looms is *how to design and improve training simulators?*

Reviewing the history of the field would teach us that the emergence of fidelity is in part due to the lack of a consistent theory of design for training systems. The growing power of computational technologies made researchers depend on realistic simulations, and this prevented attempts at creating a theory of design in the field. We used the literature to call for a theory of design for training system.

Having a theory of design can have numerous beneficial consequences. If we have a theory, the focus and constructs of the theory is clear, and as a result, we can empirically verify the theory by conducting experiments. In fact, having a theory would make it difficult to ignore certain aspects of the training simulation and accept them intuitively. Moreover, having a theory of design would open the discussion among theorists, and would pave the way towards competing theories of design. For example, if one theory is limited to a certain domain of practice, another theory can address other fields. This would not only provide a beneficial competition between theories, but also introduce multiple theories of design used in different fields. These were examples of how introducing a theory can benefit the design and improvement of training systems.

Our goal in this paper was to show the diminishing faith in the traditional theory of fidelity, and the need to have a theory of design. We did not offer a replacement to fidelity or a theory of design. Rather, we hope to inform readers of the current state of the field to entice future theoretical works and research in proposing new theories and constructs.

Acknowledgements
This work was supported by contract ONR # N00014-18-C-7015 funded through and with Charles River Analytics Inc. Helpful comments from Ashley McDermott on the earlier drafts of this paper is gratefully acknowledged.

Disclaimer
Frank E. Ritter is required by Pennsylvania State University to include this paragraph [sic]: “Frank E. Ritter, the co-author of this paper, have financial interest with Charles River Analytics Inc.; a company in which Frank E. Ritter provides consulting services and could potentially benefit from by the results of this research. The interest has been reviewed and is being monitored by the Pennsylvania State University in accordance with its individual Conflict of Interest policy, for the purpose of maintaining the objectivity of research at the Pennsylvania State University.”

References

1. Alluisi, E. A. (1991). The development of technology for collective training: SIMNET, a case history. *Human Factors, 33*(3), 343–362.
2. Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine, 89*(3), 387–392.
3. Havinghurst, L. C., Fields, L. E., & Fields, C. L. (2003). High versus low fidelity simulations: Does the type of format affect candidates’ performance or perceptions? In *27th Annual IPMAAC Conference on Personnel Assessment: Exploring New Horizons in Assessment.*
4. Hays, R. T. (1980). *Simulator fidelity: A concept paper* (ARI Report No. 490). Alexandria, VA.: U.S. Army Research Institute for the Behavioral and Social Sciences.
5. Allen, J. A., Hays, R. T., & Buffardi, L. C. (1986). Maintenance training simulator fidelity and individual differences in transfer of training. *Human Factors, 28*(5), 497–509.
6. Miller, R. B. (1954). *Psychological considerations in the design of training equipment* (WADC Report No. 54-563, AD 71202). Springfield, OH: Carpenter Litho & Prtg. Co.
7. Mirabella, A., & Wheaton, G. R. (1974). *Effects of task index variations on transfer of training criteria* (Report No. NAVTRAEQUIPCEN 72-C-0126-1). Orlando, FL: Navy Training Equipment Center.

8. Matheny, W. G. (1978). The concept of performance equivalence in training systems. In D. E. Erwin (Eds.), *Psychological Fidelity in Simulated Work Environments*. Toronto, Canada: Proceedings of Symposium at the Annual Meeting of the American Psychological Association.
9. Kinkade, R. G., & Wheaton, G. R. (1972). Training device design. In H. P. VanCott & R. G. Kinkade (Eds.), *Human Engineering Guide to Equipment Design*. Washington, D.C.: American Institutes for Research.
10. Fink, C., & Shriver, E. (1978). *Simulators for maintenance training: Some issues, problems and areas for future research* (Report No. AFHRL-TR-78-27). Lowery Air Force Base, CO: Air Force Human Resources Laboratory.
11. Hays, R. T., & Singer, M. J. (1989). *Simulation fidelity in training system design: Bridging the gap between reality and training*. New York: Edwards.
12. Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction, 15*(2), 40–47.
13. Cant, R., & Cooper, S. (2010). Simulation-based learning in nurse education: A systematic review. *Journal of Advanced Nursing. 66*(1), 3–15.
14. Blaiwes, A. S., & Regan, J. J. (1986). Training devices: Concepts and Progress. In J. A. Ellis (Ed.), *Military Contributions to Instructional Technology*. New York: Praeger Publishers.
15. Gagne, R. M. (1954). Training Devices and Simulators: Some Research Issues. *American Psychologist, 9*(3), 95–107.
16. Jones, E. R., Hennessy, R. T., & Deutsch, S. (Eds.). (1985). *Human Factors Aspects of Simulation*. Washington, DC: National Academy Press.
17. Alessi, S. M., & Trollip, S. R. (1991). *Computer-based instruction: Methods and development.* Englewood Cliffs, NJ: Prentice Hall.
18. Flach, J., Hancock, P., Caird, J., & Vicente, K. (Eds.). (1995). *Global perspectives on the ecology of human-machine systems*. Hillsdale, NJ: Lawrence Erlbaum.
19. Parrish, R. V., McKissick, B. T., & Ashworth, B. R. (1983). *Comparison of Simulator Fidelity Model Predictions with In-simulator Evaluation Data*. Technical Paper 2106, Hampton, VA: NASA Langley Research Center.
20. Noble, C. (2002). The relationship between fidelity and learning in aviation training and assessment. *Journal of Air Transportation, 7,* 34–54.
21. Lewis, R., Strachan, A., & Smith, M. M. (2012). Is high fidelity simulation the most effective method for the development of non-technical skills in nursing? A review of the current evidence. *Open Nursing Journal, 6,* 82–89.
22. Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P., & Hamstra, S. (2011). Technology-Enhanced simulation for health professions education: A systematic review and meta-analysis. *Journal of American Medical Association, 306*, 978–988.
23. Fink, C., & Shriver, E. (1978). *Simulators for maintenance training: Some issues, problems and areas for future research* (Report No. AFHRL-TR-78-27). Lowery Air Force Base, CO: Air Force Human Resources Laboratory.
24. Swezey, R. W., Perez, R. S., & Allen, J. A. (1991). Effects of instructional strategy and motion presentation conditions on the acquisition and transfer of electromechanical troubleshooting skill. *Human Factors, 33*, 309–323.
25. Beaubien, J. M., & Baker, D. P. (2004). The use of simulation for training teamwork skills in health care: How low can you go? *Quality and Safety in Health Care, 13*(1), 51–56.
26. Smallman, H. S., & St. John, M. (2005). Naïve Realism: Misplaced faith in the utility of realistic displays. *Ergonomics in Design, 13*(Summer), 6–13.
27. Çöltekin, A., Francelet, R., Richter, K., Thoresen, J., & Fabrikant, S. I. (2018). The effects of visual realism, spatial abilities, and competition on performance in map-based route learning in men. *Cartography and Geographic Information Science, 45*, 339–353.
28. Dahlstrom, N., Dekker, S., van Winsen, R., & Nyce, J. (2009). Fidelity and validity of simulator training. *Theoretical Issues in Ergonomics Science, 10*(4), 305–314.
29. Cox, J. A., Wood, R., Boren, L., & Thorne, H. W. (1965). *Functional and appearance fidelity of training devices for fixed procedures tasks* (HumRRO Report No. 65-4). Alexandria, VA: Human Resources Research Organization.

30. Grimsley, D. L. (1969). *Acquisition, retention and retraining: Effects of high and low fidelity in training devices* (Report No. 69-1). Alexandria, VA: U.S. Army Human Resources Research Organization.
31. Perfect, P., Timson, E., White, M. D, Padfield, G. D., Erdos, R., & Gubbels, A. W. (2014). A rating scale for the subjective assessment of simulation fidelity. *Aeronautical Journal, 11*(1206), 953–974.
32. Ragan, E. D., Bowman, D. A., Kopper, R., Stinson, C., Scerbo, S., & McMahan, R. P. (2015). Effects of Field of View and Visual Complexity on Virtual Reality Training Effectiveness for a Visual Scanning Task. *IEEE Transactions on Visualization and Computer Graphics, 21*(7), 794–807.
33. Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., Orvis, K. L., & Conkey, C. (2009). Relationships between game attributes and learning outcomes: Review and research proposals. *Simulation & Gaming, 40*, 217–266.