Failed Games:
Lessons Learned from Promising but Problematic Game Prototypes in Designing for Diversity

Max Seidman, Mary Flanagan, Geoff Kaufman
Tiltfactor Laboratory, Dartmouth College
245 Black Family Visual Arts Center
Hanover NH 03755 USA
603-646-1007
max.seidman@dartmouth.edu; mary.flanagan@dartmouth.edu;
geoff.kaufman@dartmouth.edu

ABSTRACT

Iterative game design approaches have proven effective in creating persuasive games, but these approaches inevitably lead to as many abandoned designs as ones that are pursued to completion. This paper serves as a reflective and instructive post mortem for the unpublished non-digital game prototypes developed for our team's "Transforming STEM for Women and Girls: Reworking Stereotypes & Bias" (BIAS) research project. We outline three abandoned designs and explain why they were ultimately not pursued, focusing on the challenges of balancing enjoyability, feasibility of production, and impact. We discuss design strategies, including: masking games' persuasive intentions, prioritizing prototypes with their efficacy-to-cost ratio in mind, and designing for fun first. This discussion offers insights into the design of both non-digital and digital "games for impact" that allow designers and researchers alike to learn from these promising but problematic prototypes.

Keywords
Iterative design, game design, persuasive games, gender stereotypes, STEM

INTRODUCTION
The iterative design method allows game designers to create, test, and evaluate multiple problem-solving approaches quickly and efficiently (Fullerton et al. 2004). Such a method is particularly vital when attempting to tackle new and difficult problems with games, for which no best practices or little foundational research exists. Iterative design calls for designers to create many rough prototypes as quickly as possible (stressing quantity over quality), and then to pursue those prototypes that are deemed to hold the most promise.

The most promising of prototypes necessarily receive the most attention, but we can also learn a great deal from the less promising, abandoned designs, the 'unsung sacrificed heroes' of the iterative design process. Designers and researchers alike often avoid closely studying or analyzing their own failures, and the study of failure is an under-utilized aspect of reflective practice than can contribute substantially to the growth and maturity of the field of game design.

Starting in 2011, in our lab's "Transforming STEM for Women and Girls: Reworking Stereotypes & Bias" research project (hereafter referred to as the "BIAS project"), we aimed to design and study game-based interventions to reduce biases and encourage broader participation for women in science, technology, engineering, and math (STEM) domains. Because our prior work has revealed that games themselves are powerful sites for enculturation (Flanagan 2009), and that values are often embedded into games through a variety of game elements (Flanagan and Nissenbaum 2015), we approached our work with biases and stereotypes thoughtfully and carefully, with a priority on evidence-based design. Our team developed, tested, researched, and produced a suite of games, both digital and non-digital, that would be accessible and implementable in a variety of school, after-school, and home settings. The target audience for these games was to be primarily middle school and high school students in the United States (ages 11-15). In order to maximize the impact of these interventions, the games were playtested with, and created for, diverse audiences. Our team included players from differing socio-economic strata, races, ages, and ethnicities, in order to help ensure that our designs would address issues of self-efficacy and interest in STEM for students across a broad range of underrepresented social categories.

RATIONALE

Much research has pointed to the specific ways in which girls (and underrepresented populations in general) might further succeed in STEM areas in the US if particular social and cultural aspects were to improve (Hill et al. 2010). Some of these barriers might be changed through well-designed, collaborative game activities that can engage and energize players around these issues. For example, games that address psychological obstacles to identification with STEM pursuits may be especially effective for pre- and early-adolescent players. Pajares (1996; 2005) and others (e.g., Cvencek et al. 2011) have documented the decline in self-confidence in STEM areas that begins to emerge for girls in middle school and continues to amplify in high school and beyond. Girls consistently report less self-efficacy in STEM than do boys, despite the fact that actual gender differences in STEM performance are by most accounts insufficient to explain the lopsided participation of males and females in STEM courses and careers (Hyde et al. 2008).

With a massive review of twenty years of research regarding the lack of women in STEM (Hill et al. 2010) as a starting point, our exploratory work aimed to offer testable proof-of-concept activities centered on several distinct practical approaches and interventions for increasing representation in STEM. Among other strategies, we focused on designing game-based interventions to counteract **implicit bias** and **stereotype threat**—two particularly powerful psychological barriers that may benefit from small changes that can end up making large differences in girls' experiences. All of the games created in the course of the BIAS project were translations or variations of strategies that prior psychological research has shown to be effective for combating the effects of stereotype threat, implicit bias, and other key psychological barriers to underrepresented students in STEM.

Also known as 'hidden bias' or 'unconscious bias', implicit bias stems from the mind's natural (and automatic) means of constructing mental *schemas*, or representational templates, for understanding broad categories, including social identity groups (Greenwald and Banaji 1995; Greenwald and Farnham 2000; Kang 1995). The schemas we use to categorize the world are generally unconscious, but are good predictors of individuals' perception and behaviors (particularly less deliberative, more spontaneous

ones). In the case of the underrepresentation of girls and women in science, the implicit bias that girls' peers, families, and teachers (as well as girls themselves) hold may limit girls' self-efficacy, interest, and achievements in STEM. Implicit bias is essential to unpack because, often, one thinks of 'discrimination' as an act perpetrated knowingly by prejudiced individuals acting to cause someone harm. Implicit bias shifts this discourse to focus more on environmental or socio-cultural causes of bias that are less volitional and more unconscious and pervasive, but, at the same time, may be reduced through the use of conscious strategies intended to reverse the 'metal habit' of bias (Devine 1989).

Stereotype threat refers to the anxiety or concern that arises in situations in which a person has the potential to confirm a negative stereotype about his or her identity group (Steele and Aronson 1995). Steele and Aronson showed that even subtle reminders of culturally-shared stereotypes that predict lower aptitude or poorer performance from certain groups can disrupt the performance of an individual who identifies with that group. Across experiments, research has shown that members of groups subject to disparaging, negative stereotypes in a given domain will experience performance-debilitating anxiety unless bolstered in some way. For example, stereotype threat is known to emerge if a test is introduced to a member of a stereotype-targeted group as being diagnostic of a person's innate ability (Bell et al. 2003; Steele and Aronson 1995). Researchers have found that women who were told that a math exam had "shown gender differences in the past" scored lower than other women with equivalent math backgrounds (Johns et al. 2005). There are many such studies supporting the evidence that stereotype threat has real-world impact (Hill et al. 2010). Certain types of "strengths-focused'' activities have been found to help 'inoculate' learners against the effect. For example, research indicates that subtly shifting female students' focus onto their strengths, rather than potential weaknesses, can work against the effects of stereotype threat in STEM (Johns et al. 2005).

With this vast body of psychological research as a foundation and starting point, our team iteratively crafted over twenty game prototypes that attempted to counteract stereotypes and biases, using strategies including promoting growth mindset (i.e., the perception that abilities are not fixed, but rather amenable to change through experience and practice: Aronson et al. 2002), teaching about stereotype threat (Johns et al. 2005), utilizing techniques from embodied cognition (i.e., using bodily or visceral experiences to trigger emotional or cognitive changes: Wilson 2002), and improving spatial reasoning abilities (Sorby and Baartmans 2000).

Seven prototypes in the BIAS project reached states at which they were studied experimentally. To date, three of these games have been produced and distributed, with one additional game currently in production; two success cases are described at the end of this paper. The games that reached the production stage were deemed to meet the 'dual bottom line' of enjoyability *and* impact. The games had to be good games first, and, additionally, meet the challenge of producing quantifiable results as evidence attesting to their positive impact on players. Prototypes that failed to meet these either of these criteria were abandoned; this was the outcome for each of the abandoned prototypes described in this paper. Failures recognized early on in the iterative design process, however, often informed and contributed to the success of other prototypes.

CASE STUDIES
The abandoned prototypes to be presented were each judged to be problematic for different reasons. Through playtesting, for example, the board game *Doubtlanders* was

found to be lacking in several respects: the game's pacing was too slow, it did not include enough opportunities for meaningful choice, and was it not balanced enough to support its intended impact in overcoming stereotype threat. During development, the party game *Teaser Totter* was revealed to be impractical to develop, and, after controlled research, the board game (and possible digital game) *Skyline* was not able to be shown to be effective for producing its intended impact. Each of these failed designs provided our team with a host of new insights and helped guide the design of subsequent games that successfully balanced the criteria of being fun, practical to produce, and impactful.

Doubtlanders

The cooperative board game prototype *Doubtlanders* was developed in order to illustrate the effects of stereotype threat and to model a growth mindset in STEM. Research has shown that, in the case of stereotype threat, simply informing participants about the existence and debilitating impact of stereotype threat can mitigate its effects on self-belief and performance (Hill et al. 2010).

During gameplay, players take on the role of the fantastical citizens of the fictional 'Barony of Doubtland' and work together to combat the dictatorial rule of their evil ruler, 'Baron Nefarious'. Players take turns moving their characters around a nonlinear board (see Figure 1), defeating challenges and improving their characters' skills. The tension in the game arises from Baron Nefarious himself: in between the players' turns, the Baron moves about the board and places difficult-to-defeat 'Dastardly Scheme' cards. These cards require the players to strategize and accumulate skills in order to bolster the character's sense of self-efficacy and remove the schemes, or risk losing the game.

The crux of *Doubtlanders'* impact lies in its modeling of stereotype threat. Throughout the game, Baron Nefarious threatens the characters using stereotypes about them depicted on 'Doubt' cards; these temporarily reduce the characters' skills from their full potential and make it more difficult for characters to overcome the Baron's schemes. Players must spend time overcoming these Doubt cards in order to restore their skills.

Many games intended to promote diversity take the very literal approach of assigning the player (or non player characters) real-world minority group status, and then seek to confront existing cultural biases against those groups (e.g., the representation of an African-American graduate student in *Fair Play*: Carnes et al. 2013). *Doubtlanders* allows players to assume the roles of members of a *fictional* species, with real-world stereotypes randomly ascribed to them: for example, the Toadstool Folk are thought of as cowardly, the Banshees are thought of as evil, and so forth. The use of fiction to address real-life social biases is a tactic that has been used in other popular media, such as such as the depiction of racism against aliens in the film *District 9* (2009), and against elves in the cross-platform *Dragon Age* video game series (2009, 2014), to cite two recent examples. Researchers and game designers, however, have not fully adopted this strategy of fictionalizing biases to address a real-world problem, particularly in the domain of pervasive games or "games for change." Nonetheless, given that in related psychological research we have found that fiction is powerful and transformative means of shifting individuals' beliefs, behaviors, and identities (Kaufman and Libby 2012), we believed the use of fiction and light metaphor would be an effective and promising means of modeling the tricky dynamics of stereotype threat. Furthermore, we hypothesized that using fictional depictions of bias would increase the game's potential to reach a wider array of target demographics by not confining its appeal to a single real-life stereotyped group.
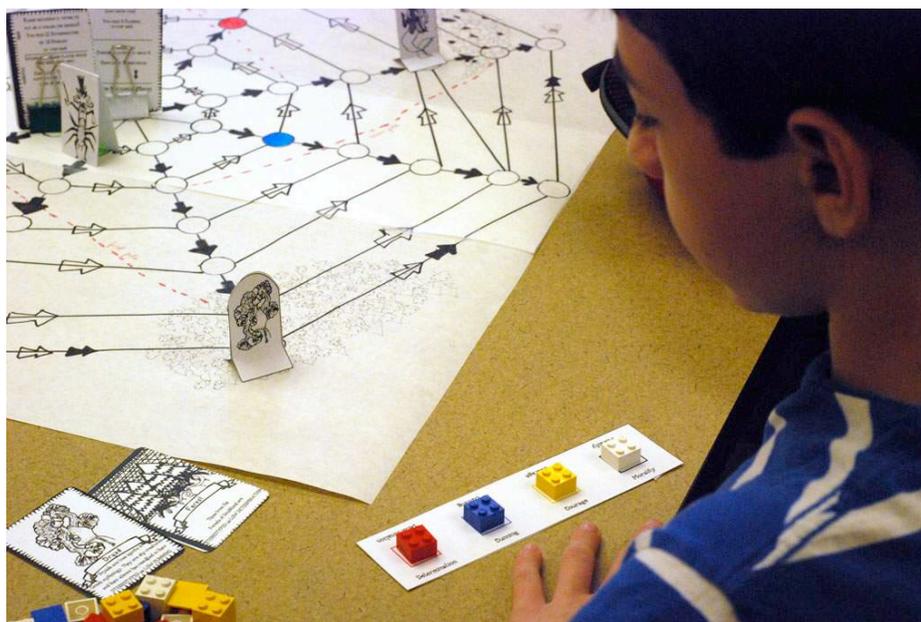
Figure 1: *Doubtlanders* being playtested by middle school students at a local public library, showing the board, the characters, Baron Nefarious, and character skills.

At the same time, while *Doubtlanders* wholly fictionalized the concrete details of bias depicted in the game (i.e., the groups involved and the stereotypes attributed to them), the game only slightly fictionalized the *processes* of experiencing and overcoming stereotype threat. Tellingly, during playtesting, it became apparent that players found these mechanics painfully clear in their intention: to illustrate the anxiety and self-doubt triggered by stereotype threat and teach methods to overcome the impact of stereotype threat on performance. The depiction of stereotype threat was quite literal.

The problem of literally representing the process of stereotype threat was inevitable. *Doubtlanders* needed to be explicit in its purpose, or else risk the possibility of players not being able to transfer and apply concepts they learned in the game to real-world examples or experiences of stereotype threat. The game relied on players grappling with their character's self-doubt and anxiety on nearly each turn. Ultimately, the direct nature of the modeling of this serious topic led to *Doubtlanders* being perceived by most players as dour and heavy in tone, instead of whimsical and fun. The prototype was eventually abandoned due to the challenges posed by the problematic prototype: simply put, the game was on-topic but not enjoyable. Addressing stereotype threat directly was a key challenge to keeping the game's tone light and fun.

When examined with the dual bottom line criteria of enjoyment and impact in mind, however, the *Doubtlanders* prototype did not meet our needs. *Doubtlanders* had potential to be effective at modeling stereotype threat, but did so at the expense of the second bottom-line criterion of enjoyability. Directly modeling issues such as stereotype threat and the means for overcoming it proved too overt. The game even used the word "stereotype," which, in a related game framing study we conducted, reduced players'

interest and blunted the impact of one of the successful prototypes discussed later in this paper (Kaufman and Flanagan 2015C; Kaufman, Flanagan, and Seidman 2015). Such overtness can ruin a game whose aim is to engage controversial or sensitive issues. A game that players do not wish to play or complete should not be made.

As our team continued development on games in our BIAS research, we were careful to learn from this attempt at literally addressing biases. For other games, we decided to address biases indirectly, a decision that led to *Buffalo: The Name Dropping Game* (described later in this paper) and, more broadly, to a design approach that further research has confirmed can increase a game's impact. When dealing with thorny issues such as biases and stereotypes, obfuscating a game's intentions while still having it facilitate effective attitude and behavior change, is not only possible – it is more effective than making the game's intentions overt (Kaufman and Flanagan 2015B, 2015C; Kaufman, Flanagan, and Seidman 2015).

Teaser Totter

The party game prototype *Teaser Totter* was developed to have two related effects: first, to lead players to confront their own implicit biases, and, second, to model the hindering effects of those biases. In the game, two teams of players pose riddles to one another, with a representative player from the "guessing" team designated to give answers in front of the entire group. Riddle Cards, consisting of five clues and an answer, are to be drawn by one team who pose the riddle to the other team.

The Riddle Card text is intentionally composed in order to defy common, everyday gender stereotypes (see Figure 2). Thus, in order to provide a correct answer, players must ideally "check" (i.e., override) any stereotypical assumptions inherent in the guesses that come to mind most quickly, such as the notion that women are not likely to be construction workers. In addition to challenging traditional gender roles, the riddles were made more difficult by allowing the opposing team posing the riddle to read the clues in any order they wished.

To illustrate the effects of implicit bias on perceivers, *Teaser Totter* attempted to use the body to physically model the mental "contortion" that implicit bias entails: that is, the ways that implicit bias skews one's judgments and perceptions, often in an unrecognized and unconscious fashion. In order to simulate this experience for players, *Teaser Totter* provides a physical manifestation of this contortion: every time a player fails to correctly guess a riddle's answer after a given clue, she must balance a physical object on her body (see Figure 3). If at any point one of the balanced objects falls, that player's team loses their chance to answer the riddle (and, thus, to gain points for that round).
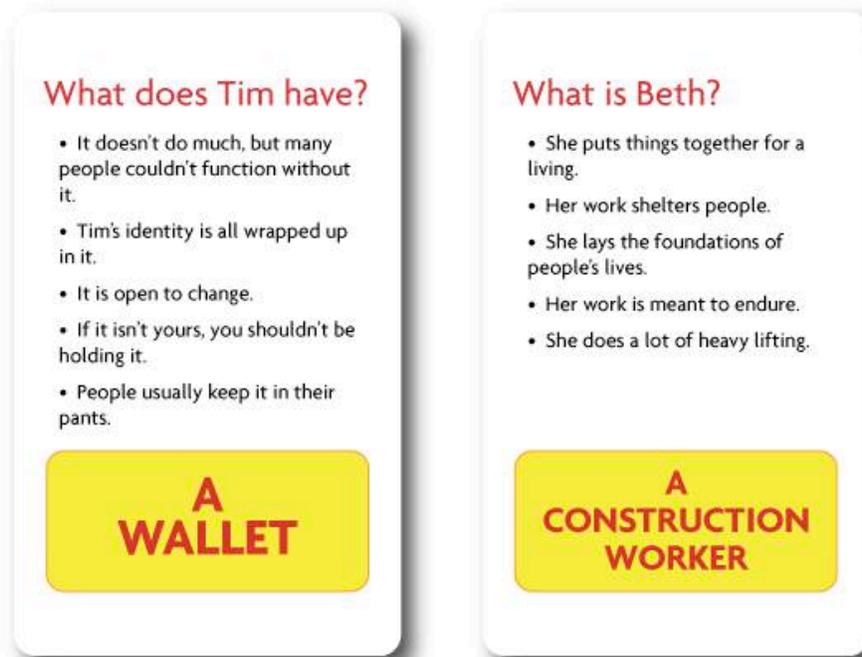
Figure 2: *Teaser Totter* riddles, including one neutral (left) and one counterstereotypical (right) answer.

Figure 3: Players balancing objects on their bodies while playing *Teaser Totter*.

Playtests of *Teaser Totter* revealed that the game was a spectacle both to watch and play, as players contorted their bodies to hold strange objects, and the riddles were challenging brain teasers to solve (even without balancing objects on one's body). The employment of embodiment (in particular, the use of embodied elements as metaphors for mental processes) has been underutilized in games for impact in general, and specifically in games dealing with biases and stereotypes. *Teaser Totter* was a promising prototype

because, unlike *Doubtlanders*, initial playtests made it clear that the game was fun, and its mechanics had the potential to be effective at combating stereotype threat.

Despite its potential for success, during development it quickly became clear that *Teaser Totter* would be challenging to produce affordably. The game's design relied a great deal on content (riddle) generation, while the game's core strengths lay in its embodied elements. The nature of *Teaser Totter*'s gameplay dictated that each riddle in the game could only be used once, necessitating a very large number of riddles, which, our team discovered, were non-trivial to write. The daunting task of writing hundreds of equally-challenging riddles that could cleverly address bias in nonobvious ways was the practical consideration that led to *Teaser Totter*'s abandonment.

It is always tempting when creating games for change to design heavily content-driven games, relying on designer-written anecdotes, situations, trivia questions, or (as in this case) riddles to change players' attitudes and behaviors. When not crafted carefully, content-driven games run the risk of having low replayability (which is especially problematic in tabletop games), while requiring high design investment. Successful content-driven tabletop games tend to fall into two categories. "Combinational" games such as *Cards Against Humanity* make use of a fairly limited collection of cards with lengthy content, but this limited collection of content is lent replayability by the combining of the content in many ways (perhaps most notably in *Cards Against Humanity*'s fill-in-multiple-blanks mechanic). "Massive Content" games such as *Taboo* and *Pictionary* rely on a substantial collection of content, of which each piece can only be used once (or once in a while), but take relatively little thought, time, or effort to develop.

	Difficult to write content	**Quick to write content**
One-time-use content	Example: *Teaser Totter*, Prognosis: Inefficient	"Massive Content" Games, Examples: *Taboo, Pictionary*, Prognosis: Efficient
Reusable content	"Combinational" Games, Examples: *Cards Against Humanity, Awkward Moment*, Prognosis: Efficient	Example: *Buffalo*, Prognosis: Very Efficient

Table 1: Content Types and Efficiency of Generation in Content-Driven Games

More broadly, it is crucial to remember that iterative game design for social impact encourages prioritization of prototypes not only by their fun and likely effects, but also (as in all disciplines) by practical production concerns. This is not to say that designers should not endeavor to realize technically challenging prototypes, but instead to let efficiency of impact play a role in the prioritization of which prototypes to bring to fruition. Ultimately, *Teaser Totter* was deemed to be too inefficient and expensive to produce, and our team questioned its efficacy-to-cost ratio.

Skyline
The cooperative spatial puzzle game prototype *Skyline* was developed to improve ten to fourteen year-olds' spatial reasoning and three-dimensional mental rotation abilities. Greater spatial reasoning skills has been shown to correlate with increased retention in

college engineering courses (Sorby and Baartmans 2000). This finding poses a particular barrier to women, who have been shown to enter college with lower spatial reasoning ability than their male peers (Linn and Petersen 1985; Voyer et al. 1995). This difference, which most scholars believe can be attributed to less frequent use of spatial toys among women during childhood rather than an inherent biologically-determined difference in aptitude between genders, can fortunately be overcome with a relatively small amount of training (Sorby and Baartmans 2000). *Skyline* was developed to help provide this practice and to enhance players' spatial skills.

Skyline challenges four players to construct a city from blocks on a square board placed between them. Each player is tasked with constructing a particular pattern on one of the sides of the board, and the players win when each side matches its player's pattern. As there are never enough blocks for each player to build her pattern independently of the others, *Skyline* was designed to force its players to cooperate and to envision their allies' patterns to succeed in the game: players must choose blocks which, when oriented correctly, help construct multiple players' patterns (see Figure 4).
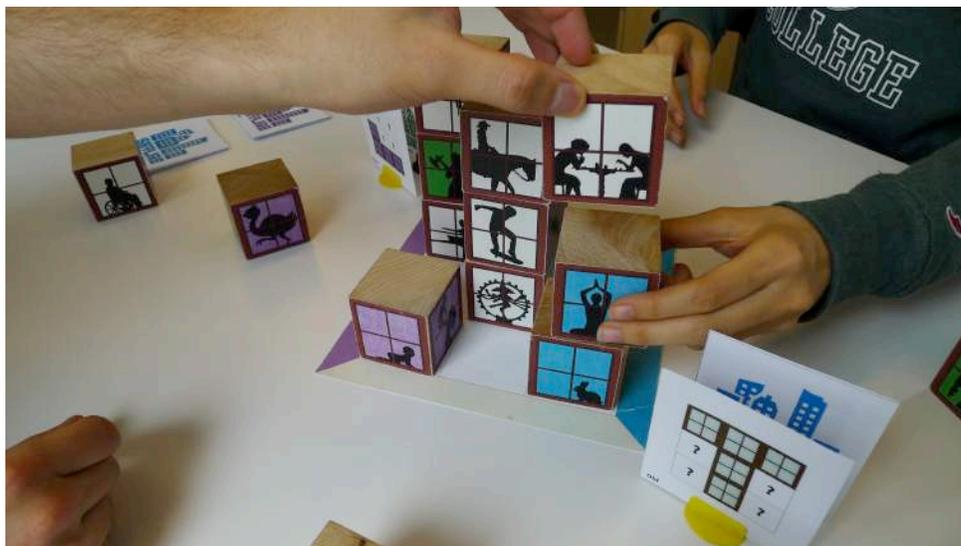

Figure 4: Two players play *Skyline*. Each player has two
cards and controls two sides of the city.

Skyline's strengths as a game lay in its engaging cooperative and communicative gameplay, the ability to easily and quickly create new puzzles for the game, and its indirect approach to combating gender bias in STEM fields (none of *Skyline*'s players guessed the game's true purpose of increasing their retention in science or engineering courses). *Skyline* was a particularly promising prototype due to its broad and diverse appeal. It was *not* a tabletop strategy game designed for the stereotypical gamer but, rather, an accessible social game that would transcend traditional player boundaries.

Our design process very consciously puts empirical research (and demonstrated evidence of players' experiences and the game's impact) at the very beginning, middle, and end of every project. At the start of a project, research guides and informs the initial game concept brainstorms. Along the way, pilot studies point to whether the team's ideas are having impact. At the end of a project, research verifies and optimizes the games' efficacy, and allows us to draw general lessons from the games to inform future

designs. Simultaneously, our games are iteratively designed with player enjoyment as the primary concern. This parallel method ensures that a game that has been developed to a production stage is going to be both fun and effective.

Unfortunately, controlled studies conducted on prototypes of *Skyline* showed no significant increase in middle school students' spatial reasoning performance over baseline scores exhibited by a no-game control condition after a single play session. *Skyline* serves as a testament to iterative design and the 'fun-first' design methodology in game design for impact. It works well as a game, but does not (at the time of this paper) produce the results we aim for. Further work will test whether the reliance on just one play session (versus repeated play experience) or some aspect of the game's design can best account for the game's lack of demonstrated efficacy. Designing for player enjoyment alongside empirically-verified effectiveness promises that, even should the worst happen and the game proves ineffectual at changing attitudes or instilling new skills, the process will result in at least an enjoyable game. This is by far preferable to inverting the process and being left with an effective but unenjoyable intervention that users simply will not engage with of their own volition.

SUCCESS CASES

We have reviewed three failed games. At this point it is relevant to describe the most successful games produced as part of our BIAS work as counterpoints for the abandoned designs described earlier. The games we will now review in fact produced great success in our 'dual bottom line' of enjoyability and impact; 'enjoyability' being measured both by extensive playtesting throughout every stage of our design process as well as subsequent purchase and distribution figures, and 'impact' measured through controlled research studies obtaining quantifiable results attesting to the games' efficacy.

Awkward Moment

Falling within the card-based party game genre alongside the popular *Apples to Apples* and *Cards Against Humanity*, *Awkward Moment* (2012) is a party card game in which players (typically aged 10-14) are challenged to react to awkward middle-school social situations. Some of these "awkward moments" are "on-topic" and include gender bias scenarios related to STEM fields (e.g., "The math team is 100% boys"), but other moments are "off-topic" and present neutral embarrassing moments unrelated to gender bias in STEM (e.g., "You sit on ketchup at lunch."). To play, players select one "Moment" card from the deck to present the social situation, to which the players must respond from their hand of 'Reaction' cards. Reactions range from silly to serious. A deck of 'Decider' cards list the criteria (such as "Most Serious") that one player uses to select the winning Reaction each round.

Via a series of controlled experimental studies, we tested the impact of *Awkward Moment* on players' likelihood of associating women and science, levels of assertiveness in responding to imagined occurrences of bias, and perspective-taking abilities, using a variety of measures. Results have shown that the game exerts a statistically significant positive effect on all of these outcomes for both youth and adult players immediately following gameplay. Importantly, these outcomes only emerge for versions of the game utilizing the "intermixing" method of balancing bias-related Moments with lighter, more whimsical Moments (Kaufman and Flanagan 2015A, 2015C). In this way, the game successfully balances the dual criteria of enjoyability and effectiveness by utilizing a less overt approach than that used by *Doubtlanders* to engage players with the serious issues of biases and stereotypes. At the same time, as discussed earlier, the fact that *Awkward*

Moment is "combinational" – the content for the Moment, Reaction, and Decider cards (while being difficult to write) is reusable by virtue of new pairings with each reshuffling of the decks – made it a relatively efficient game to produce compared to *Teaser Totter*.

Buffalo: The Name Dropping Game

A second party card game that we developed, called *Buffalo: The Name Dropping Game* (2012), is a free-form trivia game designed for players ages 14 and above. *Buffalo* is played with two card decks (one with cards listing individual adjectives and one listing nouns), and players must race to shout out the name of a real person or fictional character who is described by the pairing of adjective and noun shown when one card from each deck is revealed. For example, if the revealed cards read "British" and "Wizard," the first player to come up with a response such as "Harry Potter" or "Merlin" wins the round and claims the cards.

The key goal of *Buffalo* is to expose players to a plethora of cross-cutting categories (through the pairing of a wide array of adjectives with numerous social categories, such as race, gender, nationality, religion, profession, and ideological orientation) as a means of reducing prejudice. A series of controlled studies involving *Buffalo* revealed that the game significantly lowered adult players' prejudices and increased their concern about being biased, as measured by the Universal Orientation Scale, a measure of universal non-prejudice assessing individuals' likelihood of focusing on interpersonal similarities rather than differences, and the Internal and External Motivation to Control Prejudice scales (Kaufman and Flanagan 2015C). At the same time, as both our playtest interviews and controlled research revealed, in contrast to *Doubtlanders* very few players realized the underlying intention of the game or its connection to prejudice (instead connecting the game more to knowledge of pop culture or historical trivia). The fact that *Buffalo* had measurable impact and provided a pleasurable experience for most players meant that the game satisfied our dual bottom line criteria of fun and efficacy. *Buffalo* achieved a gold standard of efficacy-to-cost by, unlike *Teaser Totter*, featuring reusable content (stemming from the game's 200+ noun and adjective cards in combination) that was extremely efficient for the design team to produce.

CONCLUSION

Designers can learn much from their successes, and perhaps even more from their failures. Our team finds that contrasting successes and failures provides deeper insight. Comparing the elements that led to the success of *Buffalo* and *Awkward Moment* with the elements that contributed to the abandonment of *Doubtlanders*, *Teaser Totter*, and *Skyline*, yielded generalizable lessons relevant to any designer creating games for impact.

Just as with any game, when creating games for impact, designers must focus on the player experience *first* but, at the same time, negotiate the additional responsibility of verifying the game's intended impact on its players. As our work has shown, persuasive games can be both more fun *and* more persuasive (especially when dealing with controversial issues) when their messages are less obvious, or even masked, in gameplay. Furthermore, when possible, content-driven games should strive to be designed with reusable and/or combinatorial content; otherwise they risk being extremely challenging for small teams to produce well and to capture audiences for repeated play experiences. The case studies and reflections offered here provide designers and researchers of games for impact with three concrete standards (and the means to assess them) to forecast the ultimate success of their prototypes.

ACKNOWLEDGMENTS

BIBLIOGRAPHY

Aronson, J., Fried, C. B., and Good, C. (2002) "Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence," in Journal of Experimental Social Psychology, 38, pp. 113-125.

Bell, A.E., Spencer, S.J., Iserman, E. and Logel, C.E.R. (2003) "Stereotype Threat and Women's Performance in Engineering," in Journal of Engineering Education, 92(4), pp. 307-312.

Cards Against Humanity LLC. (2011) Cards Against Humanity [Tabletop Game], Cards Against Humanity LLC, USA: played 11 December 2014.

Carnes et al. (2011) Fair Play [PC Computer], University of Wisconsin-Madison, Madison USA: played 5 January 2015.

Cvencek, D., Meltzoff, A. N., and Greenwald, A. G. (2011). "Math–gender stereotypes in elementary school children," in Child Development, 82(3), pp. 766-779.

Devine, P. G. (1989). "Stereotypes and Prejudice: Their Automatic and Controlled Components," in Journal of Personality and Social Psychology, 56(1), pp. 5-18.

Flanagan, M. (2009). Critical Play: Radical Game Design. Cambridge, MA: The MIT Press.

Flanagan, M. and Nissenbaum, H. (2014). Values at Play in Digital Games. Cambridge, MA: The MIT Press.

Fullerton, T. (2014). Game Design Workshop: Designing, Prototyping and Playtesting Games, 3rd Edition. CMP Books.

Greenwald, A. G., and Banaji, M. R. (1995) "Implicit Social Cognition: Attitudes, Self-esteem, and Stereotypes," in Psychological Review, 102(1), pp. 4-27.

Greenwald, A. G., and Farnham, S. D. (2000) "Using the Implicit Association Test to measure self-esteem and self-concept," in Journal of Personality and Social Psychology, 79(6), pp. 1022-1038.

Hill, C., Corbett, C. and St. Rose, A. (2010). Why So Few? Women in Science, Technology, Engineering and Mathematics. AAUW, Washington, DC.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., and Williams, C. C. (2008). "Gender similarities characterize math performance," in Science, 321(5888), pp. 94-495.

Johns, M., Schmader, T. and Martens, A. (2005) "Knowing is Half the Battle: Teaching Stereotype Threat as a Means of Improving Women's Math Performance," in Psychological Science, 16(3), pp. 175-179.

Kang, J. (1995) "Trojan Horses of Race (Implicit Bias Paper)," in Harvard Law Review 118(5), March, pp. 1489 – 1593.

Kaufman, G., Flanagan, M., and Seidman, M. (2015). "Creating Stealth Game Interventions for Attitude and Behavior Change: An "Embedded Design" Model. Proceedings of DiGRA 2015: Diversity of Play: Games – Cultures – Identities. (Lüneburg, Germany, May 2015).

Kaufman, G. F. and Flanagan, M. (2015A). "How Awkward! Shifting Workplace Biases with a Social Game." Under Review.

Kaufman, G. F. and Flanagan, M. (2015B). "How STEM Role Models Shift Thinking About STEM Careers in a Time Travel Game." Under Review.

Kaufman, G., and Flanagan, M. (2015C). "A Psychologically "Embedded" Approach to Designing Games for Prosocial Causes." Under Review.

Kaufman, G. F., and Libby, L. K. (2012). "Changing Beliefs and Behavior through Experience-taking," in Journal of Personality and Social Psychology, 103(1), pp. 1-16.

Linn, M. C., and Petersen, A. C. (1985) "Emergence and characterization of sex differences in spatial ability: A meta-analysis," in Child Development, 56(6), pp. 1479–98.

Milton Bradley Company (1989) Taboo [Tabletop Game], Milton Bradley Company, East Longmeadow USA: played 24 November 2010.

Pajares, F. (1996). "Self-efficacy beliefs in academic settings," in Review of Educational Research, 66(4), pp. 543-578.

Pajares, F. (2005). Gender Differences in Mathematics Self-Efficacy Beliefs. Cambridge University Press, Cambridge, UK.

Pictionary, Inc. (1987) Pictionary [Tabletop Game], Pictionary Inc., Seattle USA: played 12 December 2013.

Sorby, S. A., and Baartmans, B. J. (2000) "The development and assessment of a course for enhancing the 3-D spatial visualization skills of first year engineering students," in Journal of Engineering Education, 89(3), pp. 301–07.

Steele, C. and Aronson, J. (1995) "Stereotype Threat and the Intellectual Test Performance of African Americans," in Journal of Personality and Social Psychology, 69(5), pp. 797-811.

Voyer, D., Voyer, S., and Bryden, M. P. (1995) "Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables," in Psychological Bulletin, 117(2), pp. 250–70.

Wilson, M. (2002) "Six views of embodied cognition," in Psychonomic Bulletin & Review, 9(4), pp. 625-636.