

Alternate Reality Games For Behavioral and Social Science Research

Copyright © by
Ruthanna Gordon, Ph.D.
and ETC Press 2015

ISBN (Print): 978-1-312-78105-4
ISSN (Online): 978-1-312-91112-3

TEXT: The text of this work is licensed under a Creative Commons
Attribution-NonCommercial-NonDerivative 2.5 License
(<http://creativecommons.org/licenses/by-nc-nd/2.5/>)



IMAGES: All images appearing in this work are property
of the respective copyright owners,
and are not released into the Creative Commons.
The respective owners reserve all rights.

Alternate Reality Games For Behavioral and Social Science Research

Ruthanna Gordon, Ph.D.

Production of this book was funded by the IARPA
Office of Smart Collection TRUST Program, but it is not officially endorsed by
ODNI, IARPA, or U.S. Government. All opinions and discussions in this review
are those of the author, and do not necessarily reflect those of the ODNI,
IARPA, or U.S. Government.

Table of Contents

Executive Summary.....	1
1 Introduction.....	3
2 Current Practices and Limitations.....	7
3 Alternate Reality Games.....	15
4 What Can We Learn From Related Activities?.....	23
5 What Can We Learn From Analogous Activities?.....	31
6 How are ARGs Used in Research?.....	35
7 What Can We Learn From Other Applications of ARGs?.....	43
8 What Do Research ARGs Need For Success?.....	51
9 Making a Difference: How ARG-Based Methodologies Can Improve BSSR Research.....	87
10 Conclusions.....	97
Bibliography.....	99
Appendix 1: Games Discussed.....	110
Appendix 2: Recommended Readings.....	114
Appendix 3: What a Research ARG Might Look Like.....	116
Acknowledgements.....	119
About the Author.....	121

Executive Summary

Researchers in the behavioral, social, and neurophysiological sciences have recently begun to explore the limitations of many traditional lab and field-based studies. These include concerns over ecological validity, inconsistent replication, and recruitment of sufficiently large and diverse sample populations.

Alternate reality games (ARGs) may provide a novel research methodology that addresses some of these concerns. ARGs weave fictional narratives and problem-solving into everyday life. Designers use familiar tools and multimedia venues to engage players in psychologically meaningful interaction within a complex near-real-world context. Appropriately designed research ARGs could:

- Study social and psychological phenomena under field-like conditions, but with improved control over independent and confounding variables
- Gather detailed psychological, behavioral, physiological, and even neural data during complex social interactions, allowing researchers to take advantage of new analytic methods
- Recruit larger and more diverse participant pools than many traditional research studies
- Allow controlled study of complex social interactions involving large groups of participants
- Provide an opportunity for ground-truth replication of lab-based findings that are difficult to study in the field
- Provide an opportunity to study psychological phenomena that are rare or subtle, but clinically important.
- Do all of these things at a reasonable cost per participant and with a high ROI.

This book examines the potential strengths of ARG-based behavioral, social, and neurophysiological research, the challenges that remain to be overcome, and potential starting points for pilot testing these possibilities.

Chapter 1

Introduction

Over the last few years, Behavioral and Social Science Research (BSSR)¹ has entered what might be termed a rocky developmental period. New successes abound: the field has a solid foundation of well-replicated findings, and impressively effective field applications are common across an increasing number of domains. At the same time methodological and statistical concerns, and a growing awareness of the limitations of common research practices, have led researchers to question long-held assumptions.

- Ecological validity tests show that social science is notably poor at replicating real world interactions under controlled lab conditions (Mitchell, 2012). The relationship between research settings with “mundane realism” and “psychological realism” remains unclear, as do the precise criteria required for either type of realism.
- Complaints about the lack of direct replication studies (e.g., Cesario, 2014) have led to high-profile attempts to confirm ‘classic’ findings—with mixed results (Klein et al., 2014). Replication may also be an issue in social neuroscience, where reliability has recently been questioned both across and within subjects (Brandt et al., 2013).
- Statisticians have highlighted problems with standard analysis techniques, leading to passionate debates about the relative value of probability-based and Bayesian statistics (e.g., Efron, 2005;

¹I define BSSR as a deliberately broad term, including the full range of social sciences: psychology, behavioral economics, sociology, anthropology, neuroscience, etc.

Dienes, 2011). Meanwhile, the increasing availability of large, high-quality data sets encourage the development of novel and more complex analytic techniques—and the existence of those techniques makes ambitious data collection methods more worthwhile.

- Researchers, increasingly concerned with ensuring appropriate power, sometimes struggle to recruit participants in numbers that support accuracy; social neuroscience studies are particularly prone to under-powered designs (Button et al., 2013).

Mundane Versus Psychological Realism

“Mundane” or “experimental” realism refers to the similarity between any representation of the world (e.g., a game, lab study, or work of fiction) and everyday life. “Psychological realism” refers to the degree to which represented people (e.g., characters, research assistants following experimental scripts) act like real people, and the degree to which real people respond to the representation as they would to real events (Berkowitz & Donnerstein, 1982).

ARGs may be able to take advantage of both of these, even when they contain fantastical elements. While an alien invasion is not mundanely realistic, the integration of the game with everyday tools and activities means that participants work on more realistic tasks in a more realistic environment. Such a game can also have psychological realism: players may respond to the fictional aliens in a way that reflects, and permits controlled study of, typical responses to threatening outgroups.

- Concerns that most psychological studies depend on WEIRD (western, educated, industrialized, rich, and democratic) participants (Henrich, Heine, & Norenzayan, 2010) have led to a new push for truly diverse research populations, as well as questions regarding assumptions about the “universality” of many lab-based BSSR findings.

In their recent UAREHERE Request for Information, IARPA (2013) explored the possibility that Alternate Reality Games (ARGs)—in which carefully designed problem-solving opportunities are integrated with the use of real-world tools and social interactions, both mediated and face-to-face—may offer a novel way to address some of these issues.

- ARGs offer a more controlled environment than field studies, with greater opportunity for complex behavior and interaction than the lab—together these may support increased ecological validity.
- As a potential compromise between lab and field conditions, ARGs offer the opportunity to test a wider range of circumstances under which findings will replicate – or not.
- ARGs have the potential to provide rich psychological, behavioral, and physiological data sets gathered over extended time scales, taking advantage of new statistical techniques.
- ARGs engage participants more deeply than traditional studies, and provide tools for recruiting a wider range of people. They also have the potential to cost-effectively recruit much larger numbers—participation in ARGs typically ranges from a few hundred to several thousand (Dena, 2008a).

ARG-based research may be particularly appropriate for exploring psychological, behavioral, and social phenomena that require large Ns, long time periods, or complex contexts to study successfully—or for exploring areas where those things would allow a deeper understanding. Social interactions among and between large groups – particularly those

that evolve over time - have been especially challenging to study under controlled conditions.

A few proposals and pilots have begun to explore the potential of ARGs for research (e.g., Moran et al., 2013; Stokes, Watson, Fullerton, & Wiscombe, 2013). The ARG design community has responded to this work with a willingness to share its own expertise. ARG-based research is still in its infancy, and many questions remain to be answered. However, interest is growing.

Chapter 2

Current Practices and Limitations

Many researchers in the behavioral, neurological, and social sciences have begun to highlight limitations in their current methodologies, and to debate new ways of overcoming those limits. Some consider these issues to be harbingers of a full-fledged crisis of credibility if left unchecked (e.g., Ioannidis, 2012; Kahneman, 2012).

Ecological Validity. Traditionally, BSSR has taken place either using well-controlled lab studies, field observations, or mixed methods combining the two. There are many intrinsic differences between lab and field studies. Lab studies are better controlled, and lead to greater confidence that the independent variables under examination have a causal relationship with any effects found—within the limited context of the study. Field studies tend to draw on more diverse populations, and to study effects embedded in the full complexity of their natural social contexts—and that complexity provides endless confounds that limit causal inference. For as long as there have been psychological studies, there have been concerns about ecological validity: that is, the degree to which cognition and behavior in the lab reflect what humans do in their natural environments (e.g., Orne, 1962; Berkowitz & Donnerstein, 1982, Kingstone et al., 2003). This is particularly important from a practical perspective, since psychological interventions that only work in the lab may be of limited utility.

Mitchell (2012) performed what could be called a “meta-meta-analysis”—a meta-analysis of previous meta-analyses that compared lab and field findings for specific BSSR phenomena—examining this question.

Where earlier and more limited comparisons (Anderson, Lindsay, & Bushman, 1999) found reasonable levels of replication across methodologies, what Mitchell discovered was more disturbing. Out of 215 effects, 30 (14%) changed sign (i.e., a positive correlation versus a negative

Appropriate Levels of Analysis

One aspect of validity that is sometimes neglected is the choice of appropriate analytic levels. Modern BSSR research has a wealth of options in this regard:

- FMRIs, EEGs, and neurotransmitter/ hormone assays allow detailed study of neural processes
- Surveys and cognitive tasks let researchers focus on mental processes
- Observation and operational measures emphasize behavior at both the group and individual level.

Alternate reality games are capable of integrating all of these, and researchers may wish to consider which tools best match their questions.

It may often be tempting to think about these levels hierarchically, with neuroscience techniques seen as the deepest and most accurate level of analysis. However, the most meaningful answers are often found by testing at the level at which the phenomenon of interest takes place, or at which applications are likely to be developed. Other levels may provide additional insight if used in a considered fashion, and if we identify limitations regarding how far we can, or should, have confidence that results from one level of analysis translate to another.

one) between lab and field. While some BSSR subfields, such as Industrial/Organizational Psychology, do quite well, others do not. In particular, 26.3% of social psychology findings did not hold up between settings.

The weakness of laboratory research in social psychology appears related to a tendency toward small effect sizes (Mitchell, 2012). Those small effects, in turn, may well be related to the artificiality of the lab set-up, which is particularly ill-suited to measure psychological phenomena normally triggered by complex multi-person stimuli over extended periods of time. At the same time, it seems likely that these complex interactions are particularly vulnerable to confounds in the field, where the sheer number of social variables makes teasing out the effect of any specific one particularly challenging. Similar issues hold for social neuroscience research, where conditions are often less realistic and *N*s are even smaller—and field comparisons essentially impossible.

Social decision making may be particularly artificial in the context of the lab or fMRI scanner, isolated from ongoing relationships and interactions. In general, the more factors an individual is likely to take into account when acting in reality, the less likely a lab study is to produce ecologically valid psychological responses. The short time frame of lab studies adds to this artificiality, rendering unnecessary any consideration of long-term outcomes. This may have minimal relevance for studies of perceptual judgment, but may have a significant influence on processes such as trust formation or conformity.

Replication. Psychologists are also growing increasingly concerned about the minimal degree to which most BSSR findings are replicated. While a few key results are repeated and expanded, a large number of findings announced with great fanfare are rarely or never replicated (Pashler & Wagenmakers, 2012). Further, failed replication attempts are often rejected for publication, and therefore frequently never even submitted—

the dreaded ‘file drawer problem’ (Howard et al., 2009). Many findings that appear to have been replicated have, in reality, been disproven by a larger number of studies that have never been publicized. Klein et al. (2013) engaged 36 labs in an attempt to replicate 13 classic studies, of which only 10 held up. These were relatively simple studies, which could be administered briefly in survey format. Further replication attempts, of somewhat more complex effects, performed even less well, with 10 out of 14 failing to repeat earlier findings—highlighting the importance of replication efforts that span settings and methodologies (Nosek & Lakens, 2014).

Ioannidis (2005) suggests that a combination of factors lead to the publication of many false research findings. These factors include small participant samples, minimal statistical thresholds, design biases, and heavy incentives to produce statistically significant findings. This increases both the need for frequent replication attempts to filter out inaccurate findings, and for the development of more rigorous and ecologically valid methodologies.

Statistics and Big Data. The details of the debate over Bayesian versus frequentist statistics are largely beyond the scope of this work. There are many excellent resources for exploring these issues (starting with the list in Fox, 2011). In brief, frequentist statistics focus on the probability of collecting particular data, given a null hypothesis (i.e., that the world itself does not contain the pattern found in the sample); Bayesian statistics focus on the probability of a particular thing being true in the world at large, given the data you have collected. For this discussion, it is the existence of the controversy that is relevant: BSSR researchers are exploring more deeply what can be inferred from different types of data sets, re-analyzing previously settled questions, and working with a wider range of techniques that are applicable to a wider range of data sets.

These discussions are vital, because large, complex data sets are increasingly becoming available for analysis. In turn, new techniques, far beyond the capabilities of either standard frequentist or Bayesian analysis, are being developed that make these sets more informative. Methods such as machine learning can find subtle patterns involving large numbers of diverse variables, with startling results. These modeling techniques have been used for everything from DNA analysis (DTRA, 2013) to improved prediction of areas at risk for mass atrocities (USAID, 2013).

Many of these efforts have used previously existing (and underutilized) data sets, but they have also opened the door to research methods that collect data in volumes that would have previously been untenable. ARG-based studies, collecting extensive and diverse data sets from large sample populations under realistic conditions, would likely be well-placed to take advantage of these new statistical techniques—although they might also need to be cautious in choosing appropriate techniques for a given data set. Such techniques are rarely “one size fits all,” and the precise combination of analyses chosen may affect the conclusions that can be drawn from each one.

Participant Recruitment and the WEIRD Problem. While field studies draw on a wider range of populations than lab, most research focuses on a relatively limited population. These participants are WEIRD—drawn from western, educated, industrialized, rich, democratic cultures (Henrich, Heine, & Norenzayan, 2010). Indeed, a significant percentage of lab studies are performed on undergraduates taking psychology courses at American universities—such people are about 4000 times more likely to participate in social psychology experiments than people from non-western countries.

While WEIRD people have much in common with those of other backgrounds, limited sampling means that we do not yet know whether many phenomena generalize. Further, many important phenomena related to

social interaction are known either not to generalize, or to be vulnerable to culture-specific interpretations of tasks and instructions. For example, in the Ultimatum Game paradigm, one of a pair of participants is given money, and offers the other some portion of it. The second partner then decides whether to accept or reject; if they reject, neither partner gets anything. WEIRD participants consistently reject low offers, punishing the “unfair” partner at cost to themselves—and they are even more prone to doing this than members of other cultures with similar tendencies. Furthermore, these tendencies are not universal: participants from some cultures will accept any split (more rational according to classical economics), and still others will reject offers that are seen as too generous. Other domains in which WEIRD cognition can be atypical include categorization, abstract moral reasoning, and visual processing (Henrich, Heine, & Norenzayan, 2010).

Some researchers interpret these findings as reflecting genuine cultural differences in perceptions of fairness, while others see dramatically different understandings of the task requirements. Either way, this suggests fundamental gaps in our understanding of how people of different cultures negotiate, interpret minimally-defined relationships, and understand economic transactions.

The WEIRD problem is compounded by the relatively small number of participants in most behavioral lab studies. While researchers are increasingly aware of the need for appropriate power, the 30-50-person sample necessary to detect a medium-sized psychological phenomenon is unsuited for examining rare but important phenomena (e.g., defection from a social contract). Recruitment on the scale needed to study these issues can be prohibitive, both due to budgetary constraints and minimal interest on the part of potential participants. Social science experiments suffer from an additional numbers problem—while many phenomena of interest naturally occur among groups of 10s, hundreds, or thousands, getting 5 people into the lab at the same time can be a challenge. While

researchers attempt to generalize from lab findings with small groups, many phenomena are either known to work differently at different scales, or hypothesized to do so based on a strong theoretical justification (Schneier, 2012).

For all these reasons, a number of researchers have strongly suggested the use of novel quasi-experimental and mixed methods as a way to balance the strengths of both lab and field methods, compensate for the weaknesses of each, and resolve cross-setting conflicts (Grant & Wall, 2009; Fulmer & Gelfland, 2012). Alternate Reality Games represent one intriguing possibility for operationalizing these recommendations.

Alternate Reality Games

Alternate reality games (ARGs) are a form of interactive entertainment that interweaves the elements of fiction and recreation—adventure, mystery, and high-stakes problem solving—with the tools and interactions of everyday life (IGDA, 2006). ARG participants play the game—find clues, solve problems, and interact with each other and with fictional characters—using e-mail, mobile devices, phones, and in-person interaction. Games cover time scales ranging from hours to years. Play takes place in almost any sort of location, including (but not limited to) movie theaters, museums, national parks, city centers, and offices (Montola, Stenros, & Waern, 2009).

ARGs are fundamentally a form of *transmedia entertainment* (Phillips, 2012). While most traditional games take place within a “magic circle” defined by a board, a field, or a screen, ARGs are deliberately designed to expand the play space, blurring it with other activities (Montola, Stenros, & Waern, 2009). Play for the same game may include live action, virtual interaction, and solo or collaborative puzzle-solving. Clues may be planted in movies, websites, CDs, shop windows, and any other medium that the *puppetmasters* running the game can imagine. Although particular games may deliberately place limits, ARGs may be expanded:

- *Spatially*: Although some games are local, others may be international in scope. These different scales lead to differences in the types of play available, and different challenges for designers. Local games may lean more heavily on the assumption that most players can attend live events, and can draw on landmarks and features specific to

local geography. Broadly distributed games depend more on online interaction, have the potential for a much larger player pool, and can require more extensive and difficult collaboration. An international game with a large number of players can put part of a puzzle in Amsterdam and part in Austin Texas, and could expect them to be brought together within hours or days. This spatial scaling, if applied to research, has the potential to dramatically improve participant diversity over traditional BSSR experiments.

- *Temporally*: Play takes place over longer time periods than most traditional games (and lab experiments), and may integrate itself into everyday life in a way that a paused video game or a lab experiment cannot. While some ARGs involve intensive interaction over the course of a weekend or a few days, most stretch over weeks or months. This expanded time frame allows for organic development of social structures, relationships, and interactions, as players learn about each other and the game structure.
- *Socially*: It is not always obvious who is playing an ARG, who is passively playing, who is merely a bystander or spectator, and who is unaware of the ARG in the first place. This can be an advantage, increasing the feeling that the game is an organic part of the world—and therefore increasing the ecological validity of players' interactions within the game. It can also be an ethical hazard, raising questions about privacy and informed consent as observers can be drawn into game interactions, or unwittingly have their own words and actions become part of game records. Researchers, who operate under stricter constraints, may have to be more consistently cautious about this than game designers have sometimes been.

Examples of media, artifacts, and activities used for ARG play include:

- Websites: *I Love Bees* is famously accessed through an artisanal honey site. *Conspiracy for Good* includes sites for the nefarious corporation that drives the plot, including pages where players can submit 'job

applications,' and internal sites that must be "hacked" to access corporate plans. Game sites are used to create fictional organizations within the game world, provide opportunities for puzzle-solving, disseminate new information, or provide gateways (*rabbit holes*) into the game.

Classic and Current ARGs

"Further information" links for all ARGs mentioned in this work can be found in Appendix 1.

The Beast (2001): Players search for the "Sentient Machine Therapist" listed in the credits of the movie A.I., and stumble onto a murder mystery requiring extensive online and off-line puzzle-solving.

I Love Bees (2004): Players solve puzzles in order to identify specific times and locations at which pay phones will ring, revealing the next piece of an ongoing storyline.

Perplex City (2005-2007): In a modern treasure hunt, players race to find a \$100,000 prize that could be hidden anywhere in the world.

The Optimist (2013): Disney's first foray into the ARG realm starts with a blog about one girl's search for information on her grandfather, requires collaborative research into the history of the 1964 World's Fair, and inducts players into a "secret society" focused on creating a technological utopia.

Ingress (2012-present): Google's in-progress mobile phone game gets players out into the streets, gathering in groups to claim public landmarks as "portals" for the Enlightenment or the Resistance.

- Physical artifacts: Even in the internet age, physical artifacts can create an additional level of realism and engagement. These can range from an ordinary object given special game-related meaning (e.g., the jars of honey containing clues for *I Love Bees*, an “antique” record reflecting the fictional history of *The Optimist*) to fantastic devices (e.g., Momentum created “working” machines for communicating with ghosts that players could actually use—with a little tinkering to “fix” them).
- Movies and television: Many games are centered around a large media franchise, and take advantage of those central properties to plant information that’s particularly meaningful to players. *Lost* and *Heroes* both included game-relevant information in their TV shows, and built the games around the events of specific seasons. Jeanine Sala’s “Sentient Machine Therapist” line in the A.I. credits is a well-known example.
- Online videos & audio: Short videos or recordings from in-game events or characters can add another layer of realism. These can provide a forum for direct interaction with characters, or allow new instructions to be shared using the story framework. *Ingress* uses weekly broadcasts to update “agents” on game progress. *Arcane Gallery of Gadgetry* put together clever podcasts from historical figures, shared through the game’s “time machine.” *FutureCoast* made participant-created voicemails central to play, using them to provide glimpses of everyday life in possible futures affected by climate change.
- Blogs: Online interactive posts can be a primary mode of play (as in *World Without Oil*) or a way for characters to attract and interact with potential players (as in the protagonist’s Twitter feed in *The Optimist*). This type of organic recruitment contrasts with the more explicit (and frequently less far-reaching) tactics often used to gather research participants.

Another way that ARGs can blur boundaries is through the collaborative creation of the game itself. The puppetmaster(s) and their supporting team usually create the central storyline and puzzles. However, players can be apt to expand on these, often coming up with new methods of play or directions for story development that were not anticipated by the game's creators. Players also frequently create tools—by now, a process usually anticipated during design—to document play, bring new players up to speed, and improve the player community's problem-solving capacity. Creators of research ARGs may wish to take this aspect of play into consideration. ARG participants, like participants in many real life activities, may not restrict their activities to those detailed in experimenter instructions (Montola, Stenros, & Waern, 2009). While this increases ecological validity, it means that designers may wish to be cautious about the affordances and opportunities made available to their players. For example, if the speed with which players develop a specific solution is a vital dependent variable, designers might consider what alternatives a creative player might invent, and come up with ways to forestall them.

ARG play can include several broad categories of player activity. While many encourage one specific type of play, others allow different types to co-exist. Each category reflects a particular aspect of ecologically valid behavior, and different categories may be suitable for different research purposes:

Collective cognition. In collective cognition games, large numbers of players collaborate to solve problems. These games highlight the same potential taken advantage of by crowdsourcing activities (e.g., DARPA's Red Balloon challenge; DARPA, 2009). Designers of collective intelligence games must continually up the challenge level in order to maintain difficulty for thousands of international collaborators working in concert. When not prevented by the basic game design, players tend to create these collaborative groups spontaneously.

In *I Love Bees*, players worked together in online forums to decode the location of pay phones; they would then self-organize teams to answer the phones and report new clues back to the full player community (Peyton, Young, & Lutters, 2013). Players gained a collective advantage due to geographic dispersal, the wide range of expertise available, and the flexibility to collaborate in whatever combinations were most appropriate to a given challenge. Leaders emerged who were particularly skilled at facilitating these connections.

ARG-based collaboration frequently centers around online communities dedicated to the purpose—Unfiction.com, for example, is a common place for experienced ARG players to identify new games, share information, and form spontaneous collaborations. Participants in these communities, who constantly seek out new puzzles to explore, are not necessarily representative of the population as a whole. However, they may be exemplars of what large-scale collaborative groups and dispersed teams can accomplish under ideal circumstances.

ARGs involving collective intelligence activities might be used to study:

- The potential and limits of crowdsourced problem-solving
- Self-organization of spontaneous teams
- Large-scale social dynamics
- Dynamics of collective memory, perception, storytelling, and decision-making

This (along with similar suggestions below) is intended as a starter list of seed ideas, rather than a comprehensive set of suggestions. The most appropriate foci for different kinds of research ARGs remain to be determined.

Small-scale intelligence. Many ARGs design activities that encourage competition between either individual players or small groups. These

reflect everyday cognition more closely than do collective intelligence activities, and may be appealing to a broader population. Treasure hunt games, such as *Perplex City*, fall into this category. Training games often focus on this type of activity as well (Whitton, 2008; Andersen, 2011). Ingress includes activities at both levels—while some problem solving requires large-scale collaboration across international borders, stable local teams are the focus of goals such as claiming specific portals.

ARGs may be one of the few human activities in which the default is large-scale cooperation—where it is possible to apply the power of crowdsourcing, players will frequently self-organize to do so spontaneously, even if not required by the set-up. Designers who wish to encourage smaller-scale activity may seek to set up competition between players/teams, create a narrative that cues players to work in smaller groups, or create problems that do not give an advantage to larger collaborative groups. (The circumstances under which large-scale cooperation becomes a default might, in and of itself, be an interesting research topic.)

ARGs involving small-scale intelligence activities could be used to study:

- Small team dynamics (i.e., those common in offices, military squads, etc.)
- Leadership dynamics
- Trust formation and development
- Individual cognition

Story-focused activities. Not all ARG play is focused on direct problem-solving. Most ARGs have some sort of framing narrative, and in some cases players focus on taking roles in these scenarios, and exploring the implications of those roles. This type of collaborative storytelling doesn't have the same sort of clear-cut goal, but includes many types of social modeling and sense-making that are vital parts of everyday real-world behavior (Bunting, Hughes, & Hetland, 2012). These types of activities

may be useful in supporting the psychological realism of ARG-based research design.

An increasingly common form of story-focused ARG is the foresight game, intended to explore possible futures and create detailed scenarios for further exploration. *World Without Oil* is a classic example: players were asked to imagine that peak oil had been reached, and produced blog posts and other “future artifacts” detailing how their families and communities were responding to this crisis. A more formal version of this type of game, pioneered by the Institute for the Future, asks players to create “cards” that build on others’ ideas to explore a wide range of potential scenarios. The think tank uses these games to inform their predictions and recommendations; *World Without Oil* materials have also been used as a basis for classroom discussion. The Navy’s MMOWGLI effort uses this method to explore options for responding to both current concerns and hypothetical future situations. Although these games clearly involve problem solving, those efforts involve telling stories rather than finding solutions to puzzles.

ARGs involving story-focused activities might be used to study:

- Narrative comprehension and narrative influences on decision making
- Counterfactual thinking and best practices for foresight
- Social modeling and perspective-taking
- Psychological and neurophysiological effects of role-taking and empathy

What Can We Learn From Related Activities?

Alternate Reality Games take place within a larger ecology of activities involving interactive entertainment, pervasive play, and playful problem solving. The list is extensive and potentially infinitely expandable, and this work cannot provide an exhaustive taxonomy of such activities. Instead, we offer a brief exploration, and a discussion of how lessons drawn from these activities may be relevant to ARG-based research.

Larger categories

ARGs fit into several categories of activity, each of which highlights particular interesting aspects of the form.

Pervasive games minimize spatial, temporal, and/or social boundaries on play—that is, they are embedded in everyday life (Montola, Stenros, & Waern, 2009). Not all pervasive games include the types of social interaction, problem-solving, narrative complexity, or cross-media presentation that are typical of ARGs. For example, *Insectopia*, in which players collect “insects” associated with nearby Bluetooth devices, is spatially expanded but requires no social interaction and has no overarching storyline. The range of pervasive games, and the factors that contribute to their success or failure, are mark boundaries on where and when players are comfortable with different levels of play.

Interactive fiction includes any story shaped by input from participants. This includes everything from Choose Your Own Adventure books, to Massive Multiplayer Online Games, to storytelling games in which multiple players take turns adding to a narrative. Some of these forms

lack the level of complex player interaction and input that are typical of ARGs; others lack any sort of ‘puppetmaster’ role at all. However, methods drawn from all of them can help address one of the major challenges of creating a successful ARG: the creation of a coherent and entertaining narrative through improvisational collaboration.

Transmedia entertainment describes stories and games that are told across multiple media. For example, stories about Superman were first told solely in comic book format—but have since spread to television, movies, books, action figures, video games, amusement park rides, and online comic books that allow readers to choose a story’s direction. While marketing may provide the impetus for this expansion, it also reflects a new way of seeing stories. Narrative is no longer tied to a particular way of expressing it, but can be explored in whatever form seems most appropriate at any given moment—and a story can gain new facets from multiple angles of exploration. Superman gains not only new audiences, but new complexity, by taking advantage of the strengths of multiple media.

Transmedia entertainment is not solely for major media properties. Individuals now take advantage of affordable online platforms to tell stories through blog posts, text messages, Twitter, and video (Phillips, 2012). Organizations of all sizes learn to perform outreach using all these media as well. These efforts highlight not only the story or message being communicated, but the strengths and weaknesses of each medium. These may be important lessons for ARGs, which not only appear across media, but often expect players to find puzzles and plot hidden within several media. While a poster may clearly inform potential viewers that a new Superman movie is coming, the clues that you might find the rabbit hole for an ARG on that poster are more subtle. ARG creators should consider more closely exploring the strengths and constraints of various media in order to choose the best venue for any given piece of the game.

Related Forms

Geocaching is often seen as a proto-ARG. Hidden caches are shared through written clues and sets of geographical coordinates. A variety of mobile and online tools exist to help cache-hunters track down their targets. The game is relatively non-competitive, since most caches contain minor prizes or tags that give access to online information—traditionally, finders remove one item and add another of their own. Geocaching descends from the less technologically-dependent *letterboxing*, in which cache-hunters find sites based solely on written clues (Mapsurfer, 2005).

Talking About Stories

One sometimes confusing term that comes up frequently in the ARG and game-related literature is “diegetic.” It is not easily glossed with a less technical term. Something diegetic is not merely related to the story being told, but produced and/or treated as if it originated within a world where that story is true.

For example, players in *The Beast* acted diegetically when they spoke with in-game characters as if they were real people. They acted non-diegetically when they questioned whether they had been intended to guess the password for a particular e-mail address, or whether it was in fact an address for a real person that should be left alone (McGonigal, 2003). Players in many games move easily between both kinds behavior, and the distinction is key if that behavior is to be measured for research.

Both of these forms abstract out the problem-solving, puzzle-focused aspect of ARGs, as well as the geographic expansion. Participants enjoy the challenge of determining where to look for a cache—but the opportunity to see small landmarks and beautiful places that they might not otherwise notice is also valued. Like geocaching, ARGs can be used in this way to reframe, or provide new opportunities for, experience—encouraging

players to see the world in a new way. This change of perspective is one of the draws of play, but can also be used as a deliberate tool, training people to take different perspectives.

Live Action Role Playing (LARP), in contrast to geocaching, emphasizes the story-focused aspect of ARGs. Players take on the parts of specific characters who may have traits and skills very different from their own, and act out attempts to meet those characters' goals within the confines of the in-story setting (see text box). Game mechanics are used to determine whether specific actions succeed, and to resolve conflicts between characters that cannot simply be acted out (e.g., armed combat). However, the open-ended choice of actions to attempt belongs to the player/actor (Stark, 2012). While some problem-solving is likely to be involved, it is rarely at the scale or complexity typical of ARGs.LARPs also tend to be geographically and temporally focused: a particular group of players may gather one night a month in a specific park or building in order to create the latest story segment. LARPs that extend play online, or connect groups in different locations, may begin to spill into the ARG category.

The “Nordic School” of LARPs (frequently referred to as “Nordic Larp,” turning the acronym into a word in its own right) most fully illustrates the potential of this type of play (see, e.g., Nordic Larp Talks, n.d.). Nordic Larp is the literary fiction of the role-playing world, often emphasizing deep political themes and intensive perspective taking (Montola, Stenros, & Waern, 2012). While these things are not unheard of elsewhere, they are rarely deliberately encouraged or systematically organized. The *Momentum* game is a prototypical example. Over the course of a month, players took on two roles: their own everyday selves, and the ghost of a historical revolutionary possessing them. In between the events of their ordinary lives, they met to explore their characters' histories, the causes that had motivated them, and what they needed to be put to rest once more.

Like LARPs, the older practice of *epistolary games* emphasizes story. Two or more players write letters to each other in character, trying to build a narrative without any coordination other than the in-character letters. Where LARPs focus on live interaction, the letters are closer to experiences of collaborative storytelling in ARGs such as *World Without Oil*.

Training Simulations are often used for military, emergency response, and other complex skill sets that may be rarely used, but must be deployed at a high level of competence in the field. Formats may include live battle simulations, computerized virtual environments, or ARG-like combinations of online information management and live response. Interconnections between simulations and games can be extremely close, with innovations in both design and technology diffusing easily from one to the other (Smith, 2010).

Training simulations may integrate rules, goals, and feedback systems that make them somewhat game-like. However, their focus on logistical realism, on learning under extremely high-fidelity conditions, and on effective observation and after-action reviews, set them apart (Jensvald & Morin, 2004). While entertainment is not an overt goal—even a secondary one—these simulations may still be engaging simply due to the nature of the activities and situations being explored.

In comparison to other ARG-related activities, training simulations particularly stand out in their use of effective observation and review techniques. The depth and thoroughness of these techniques may provide a useful model for those seeking to create effective data collection and analysis for research ARGs.

Virtual Worlds simulate some aspects of live interaction in an online context, allowing users to manipulate (and sometimes voice) visually detailed avatars, and interact with other avatars, in real time in a compellingly represented environment. Some of the most popular worlds are games

with specific goals and long-term narrative development (e.g., *World of Warcraft*), while others simply offer a place to meet and interact with distant friends (e.g., *Second Life*). Small private-access worlds are also relatively common, and are used for a multitude of purposes including virtual meetings, training, and logistics planning.

As graphics have become more sophisticated and avatars have become more lifelike, virtual world users often have the capacity to include realistic body language and even eye gaze information in their virtual interactions. A few researchers have taken advantage of these capacities to perform psychological studies. These studies may seek a window into how online behavior differs from (or mirrors) live behavior, or may simply be looking for a way of studying large samples not limited by location. Results regarding similarities between online and live interaction have been mixed. Researchers have found consistent nonverbal social norms (e.g., eye gaze/conversational distance trade-offs) between *Second Life* and live conversations (Yee et al., 2007). However, they also showed that communication opportunities may have significantly more influence on cooperation in live interactions than in the virtual world (Greiner, Caravella, & Roth, 2014).

Bainbridge (2007) suggests that virtual worlds could be used for both (in-world) lab and field research, as well as for controlled-but-fieldlike studies—for example, of variables that influence conflict between factions. ARGs and virtual worlds have many of the same strengths with regard to BSSR research. Both attract potential research participants with compelling narratives and settings, and both facilitate complex and varied social interactions under somewhat controlled circumstances. Virtual worlds have at least one major advantage: they offer large, long-term communities whose origin and development are well-documented. ARG-based communities are (so far) shorter lived and generally smaller (the major virtual worlds have populations in the millions).

The clearest advantage of ARGs may be in their physicality—the ability to study facial expression and body language that are not under conscious control, to easily measure the physiological components of interactions, and to examine social processes fully influenced by embodiment. Embodied influences play a significant role in these processes, as do physical, geographic, and social context (Smith & Semin, 2007; Davis & Markman, 2012). In addition to improving ecological validity, this rich context may increase participants’ motivation and effort for research-related tasks.

ARGs’ pervasive nature could also be an advantage, both for recruitment/retention and ecological validity. While virtual worlds research generally depends on a previously self-selected population of players/users, ARGs often draw in first-time players—and integrate reminders of play into activities as commonplace as checking e-mail or carrying a phone. They also make it easy to study how those everyday tools and activities influence interaction and decision making.

All of these related forms provide an important perspective on the limits of ARG design. It can be easy to become hyperfocused on common design choices, assuming that they are necessary to making a good ARG. Some of these choices may be merely normative, and may not actually serve the ends of a particular designer. For example, several sources emphasize the “*This is Not a Game*” (TINAG) aesthetic of early ARGs, insisting that game material can never acknowledge its own fictional nature without undermining play (e.g., Szulborski, 2005). The perceived realism of early TINAG-style games may have been exaggerated by observer misinterpretations of players “performing” belief in the characters and events of the narrative (McGonigal, 2003). However, follow-up games sometimes neglected including even subtle cues to their nature. More recent ARGs have moved away from this aesthetic (Phillips, 2012), considering it to be a liability risk, and to have limited effectiveness as a storytelling device. This is a vital piece of flexibility for ARG-based research, which

- because it is research - will be ethically and legally bound to provide players with informed consent.

By looking at larger categories and related activities, we can pick out common characteristics that may be necessary to make any ARG-like activity function, and differences that demonstrate where research ARGs have room for flexibility. We can then pick and choose characteristics that support research goals and ethics. For example, geocaching and Nordic Larp represent extremes on a spectrum of perspective-taking, where players at one end have no options beyond “playing themselves,” while players at the other end may pretend to have entirely different personalities and motivations. Designs closer to geocaching may therefore be most useful when attempting to study player’s genuine reactions to situations, while tools from LARP may be useful when studying narrative processing, perspective taking, and tools for developing empathy.

What Can We Learn From Analogous Activities?

In thinking about possible uses of ARGs for research, it is also worth thinking about the types of activities for which they are most ecologically valid. Modern life, infused with technologically-mediated and technologically-provided information, offers a wealth of such situations. Even leaving aside this “transmediation” of everyday activities, large-scale cooperative problem solving has a long history that resonates with the types of situations typically created in ARGs.

Citizen Science encompasses a set of techniques in which researchers engage lay individuals and communities in research planning, data collection, and/or analysis of results. While none of these projects have taken the form of ARGs, many are formatted as computer games or include elements of gamification, and the types of coordination required are often similar to those used by ARG players. Success stories in citizen science may therefore provide models for how robust research activities can be integrated with engaging game play.

- On the *Old Weather* site, players follow the stories of historical sea voyages, while transcribing handwritten weather and climate logs. Participation is motivated by both the narratives and competition between ships (OldWeather, n.d.).
- The Audubon Society’s annual Christmas Bird Count, one of the oldest citizen science events, gets participants outside to compete over spotting birds, identifying species, and assessing their health. As with ARGs such as Ingress, the motivation to explore new places, meet new people, and socialize with friends is part of the appeal (Audobon, 2014).

Crowdsourcing and Competitions both stem from the idea that problem solving sometimes works best when it draws on as much brain power, and as many perspectives, as possible. Crowdsourcing projects, much like ARGs, encourage collective problem solving. For example, the Polymath project brings together dozens of experts at a time to solve the most difficult mathematical questions, while Galaxy Zoo combines perceptual judgments from thousands of laypeople to categorize galactic shapes (Ball, 2014).

Prize competitions, conversely, pit individuals and teams against each other to solve pressing problems, often for monetary awards. These are more similar to ARGs that focus on individual cognition, in that they provide an exciting motivational framework to push people towards their goals. The X-Prize for private space exploration is among the most famous, but prize competitions are used to solve problems in a wide range of fields and sectors. These contests may suggest frameworks for increasing motivation and engagement in ARGs.

Self-Organized Collaboration. The modern workplace offers many opportunities to self-organize around specific projects and work with geographically distributed teams. Particularly for skilled knowledge workers, exploration and collaboration are necessary to define tasks as well as to fulfill them. The similarity to ARG play appears substantial (Gurzick et al., 2011). Education researchers have suggested that ARG play supports “21st century literacies” required for creative and self-directed problem solving (Bonsignore, Hansen, Kraus, & Ruppel, 2012). Such skills are both broadly applicable and widely used, and may include:

- Information gathering
- Sense-making
- Project management
- Collaboration and team-building

Historic parallels to this type of activity are not difficult to find, and range from the growth of communities around early European coffee shops (and the subsequent emergence of the stock market) through the organization of underground revolutionary movements. However, in the 21st century this type of self-organization seems to have become more widely recognized, and more supported by mainstream culture. It deserves closer study, and ARGs may be a useful tool with which to study it.

How Are ARGs Used in Research?

A wide range of lab-based studies could be made to fit the broadest definition of games. According to McGonigal (2011), games ask players to achieve specific goals under a set of interesting constraints; players must receive regular feedback and participation must be voluntary. McGonigal would probably consider most lab studies to be poorly designed games, with insufficiently interesting goals and constraints. ARG-based research may be seen as a research scenario creation method that brings challenge, motivation, and complexity up to game-like, and life-like, levels.

For the purpose of this overview, we will focus on research that uses alternate reality games or ARG-like activities consciously and deliberately. We will start by looking at research *on* games: the young but rich body of work that examines the experiences and dynamics of play in ARGs and ARG-related activities. These findings, delineating the similarities and differences between in-game and real-world cognition and behavior, are highly relevant to any ARG-based research that hopes to draw conclusions generalizable outside a specific play situation. We will then give a brief overview of how pre-existing ARGs have been used as research tools, drawing lessons that may apply to more intentionally designed research ARGs. Finally, we will focus on the limited but intriguing first forays into genuine ARG-based research.

The study of ARGs and similar activities is a relatively new area of research, and has neither been claimed by any one discipline, nor (yet) formed the core of a cohesive interdisciplinary community. Instead, ARG researchers focus on questions and methods drawn from their home disciplines: most commonly ubiquitous/pervasive computing, digital game

studies, or cultural studies (Stenros, Waern, & Montola, 2012). Small clusters of research may often take place without reference to each other, and room remains for the formation of a new and valuable dialogue.

Experiences and Dynamics of ARG Play

Collective intelligence is not only a notable part of many ARGs, but an area of vital interest for many organizations (National Research Council, 2013). Where these organizations often struggle with the best ways to draw on group-based skills and knowledge, ARG structure seems to encourage the spontaneous formation of temporary teams, and distribution of tasks, appropriate to a given problem (McGonigal, 2007). *I Love Bees*, in which geographically distributed clues must be collated and analyzed, excelled in this area.

McGonigal (2007) suggests, based on the *I Love Bees* experience, that collective cognition, cooperation, and organization, are encouraged by three aspects of game design:

- Massively distributed content: the spread of information, and the geospecific location of subtasks, required swift coordination across sites.
- Meaningful ambiguity: Information, once gathered, both required and rewarded deeper exploration and comparison across sources.
- Real-time responsiveness: Available online tools facilitated swift connections between those who had gathered different pieces of information, and who had different skills appropriate to interpreting it. This swift feedback not only rewarded those willing to work collectively, but avoided the hindrances that often discourage distributed teamwork in other settings.

Gurzick et al. (2011) suggests that ARGs may provide a model for improved ad hoc team formation relative to that observed in more overtly

professional settings. The high levels of motivation and successful collaboration may capture the true limits of distributed teamwork under ideal circumstances, with team members self-selected for their willingness and ability to collaborate at a distance.

Spontaneous collective intelligence in ARG play is not equally distributed, but depends on the emergence of organizers and leaders who have both time and skill available to help direct game play. This is, in fact, another key characteristic of ARG play. Level of engagement is highly variable, ranging from a small community of deeply involved problem solvers to a large group who do little more than observe the spectacle of play (Dena, 2008b). Players may move between these tiers depending on real-life commitments, and create tools and documents to guide those moving into a more active play phase (O'hara et al., 2008). Phillips (2012) suggests that 80% of players simply observe passively, 15% interact in some fashion, and 5% are the "superfans" who spend extensive time solving puzzles and creating supplementary materials. Research designers may wish to take this distribution into account when recruiting, but may also consider including challenges and activities that will gather meaningful data from each level.

Collective action also takes place on different story levels within players. The same player is likely to act both in settings where they are asked to speak and act fully within the story world perspective, and in out-of-character forums where they strategize freely (O'hara et al., 2008).

Ultimately, multi-level play may prove to be an important psychological difference between game-play and non-game interactions. Interaction and movement between ordinary and playful life² is a key part of

² I use here Mark Twain's (1876) definition of play as "anything a body is not obliged to do." As Tom Sawyer elegantly demonstrated, activities treated as enjoyable recreation may still be productive and consequential.

the experience of ARG participation (Stenros, Waern, & Montola, 2012). This integration into everyday life gives ARGs their potential for ecological validity. However, it also comes at a cost, because everyday life is not controllable by the experimenter, but nevertheless influences in-game choices and behaviors.

This highlights another psychological difference between ludic and non-ludic life: contrasts and coincidences between the game and the everyday world are often mentioned by players as peak moments of enjoyment and entertainment during play. The experiences of simultaneously acting in these “different worlds” may be one of the unique experiences that make ARG participation appealing, even as it reduces the similarity to ordinary experience.

In general, fictional experiences provide this type of interesting contrast, in which events are treated as “partly real.” Players, readers, and viewers enjoy feeling as if an experience were real—to a point. *Narrative transportation* describes the process of being psychologically “swept up” in a story (Gerrig, 1993). Greater sensory detail, whether described in a novel or actually perceptible in the villain’s headquarters at the climax of an ARG, can increase transportation—as can increased psychological detail and opportunities for fully modeling characters as other people.

Greater transport can cause people to react to a narrative in more ecologically valid ways—more emotionally and cognitively similar to the way they would react to the same events in a more mundanely realistic setting. It may also increase the likelihood that people will learn from a narrative, treating information from it as applicable in the real world (Green & Brock, 2002). However, the bulk of transportation research has been performed using narratives that provide relatively minimal room for interaction. Books and movies allow us to sympathize with characters, but not work to change their fate; video games and other interactive media have received relatively little comparative research. Transport may also

affect willingness to act—for example, compelling narratives have been shown to increase organ donation registry sign-ups (Morgan, Movius, & Cody, 2009).

Research on virtual reality, where the emphasis is on technical development, often focuses on the related concept of *presence*. Presence describes the similarity between mediated and real-life sensory experience (Lombard & Ditton, 1997). It has been found to have similar effects to transport, and transport in turn may increase perceptions of presence (Pinchbeck & Stevens, 2005; Balakrishnan & Sundar, 2011). In games, opportunities for meaningful action and interaction can also contribute to presence and immersion (Ermi & Mayra, 2007).

For pervasive games, social interaction is likely to be a vital component of transportation and presence. ARGs may include every type of communication from text messages and posts to face-to-face interaction—with both other players and fictional *non-player characters (NPCs)*. The realism and bandwidth of these communications may affect the ease with which players collaborate, and the believability and motivational power of NPCs. Social interactions of both types may be key vehicles for examining hypotheses in ARG-based research. Ekman et al. (2012) suggests that physiological measurement could be integrated into game play in order to better understand these interactions. Appropriate narrative justifications could make it seem perfectly rational to put on a heart-rate monitor or EEG cap before heading out to a climactic meeting, or even to get a baseline fMRI scan. Preliminary work supports people's willingness to attribute mental states to fictional characters, and suggests that the attribution process may be very similar for fictional and real people (Besmann & Rios, 2012).

Pre-existing ARGs as Research Tools

While ARGs have rarely been designed specifically for research purposes, a number of researchers have taken advantage of existing ARGs to collect

data relevant to BSSR questions. Some of these studies have been intended to improve later ARG design; others have found in the ARG an opportunity to explore more generalizable issues. These post hoc studies suggest ways that pre-planned ARG-based research can increase opportunities for high-quality data collection.

Much research on pre-existing ARGs focuses on collective storytelling and problem solving. O'hara, Grian, & Williams (2008) studied participation patterns in *Meigeist*. The game's use of dedicated communications allowed them to examine records from online forums, chat rooms, and player's self-logged in-game actions. Researchers noted collective aspects of play, particularly the structures that players created to share information and keep those with lower time commitments updated and able to contribute. McGonigal (2007) and Gurzick et al. (2011) discuss similar collective cognition in *I Love Bees*.

For these and similar studies, analyses were largely retroactive, ethnographic, and qualitative. Analytic techniques for linguistic data have made impressive strides in the past few years (e.g., Volkova, Wilson, & Yarowsky, 2013; Brown, Watkins, & Greitzer, 2013), and it is likely that quantitative analysis of such data would now be more feasible and informative. However, this potential remains unexplored.

Pre-existing ARGs have been used for similar qualitative analysis around other topics, including counterfactual thinking in *Arcane Gallery of Gadgets* (Bonsignore et al., 2012), development of ethical reasoning in Urgent EVOKE (Boskik, 2011), effects of narrative on game play in Conspiracy for Good (Stenros et al., 2011), and player interpretations of designer intentions in *I Love Bees* (Young & Lutters, 2013). These studies focus on evaluation in response to game designer's goals around education, entertainment, and deeper involvement in real-world causes. All these games except for *I Love Bees* were designed at least to some extent with this type of evaluation in mind; although data collection was not a

primary goal, recording and archiving practices were intended to facilitate later analysis. These practices included mid-game interviews in *Conspiracy for Good* and classroom observation in *Arcane Gallery of Gadgetry*.

One example of a more quantitative study is the use of network analysis in *Reality Ends Here* (Stokes et al., 2013). Network analysis allows the identification of connections between individuals and groups within an organization or activity, including measures of group size, linkages between groups, centrality of individuals in creating networks, and the density of different parts of the network. *Reality Ends Here* was particularly well suited for this type of analysis, because game play included submission of artistic projects by collaborative groups, allowing for easy identification of working groups and the individuals constituting them.

Exploring ARG-Based Research

ARGs are a relatively novel narrative form, and alternatives to traditional BSSR paradigms have only recently started to be recognized as an urgent need. This, along with the knowledge- and resource-based barriers to ARG creation by traditional researchers, means that only a few have begun to explore the potential of ARG-based research.

Researchers at the University of Nottingham Mixed Reality Lab have argued for the use of ARGs (*mixed reality games* in their terminology) as a tool to study complex and emergent real-world events (Fischer et al., 2012; Fischer, Jiang, & Moran, 2012). Their major focus has been on disaster response, an area where simulations tend to be oversimplified and empirical data are often incomplete. ARGs allow participants to interact with each other in real-world environments, using simulated scenarios based on real events. By giving players access to a variety of expert systems, automated agents, and other tools, researchers could better understand the factors that improve coordination during crises.

Taking these proposals a step further, Moran et al. (2013) used a short ARG to study interactions between a team of players as they took instructions from a computerized agent in order to meet their goals. Players needed to reach key geographic points in order to gather specific items, and get them to a goal location. The computerized agents were, in fact, highly reliable sources of information. Interestingly, many players did not pick up on this reliability and ignored instructions that could have improved their performance. Although the set-up was relatively simple, and the design not fully experimental (e.g., the computerized agents didn't vary in how they expressed their instructions or in their actual reliability), the study gives a successful proof-of-concept demonstration for the use of ARGs to research ecologically meaningful questions.

A more recent follow-up study builds on this work to examine team coordination dynamics in a 30-minute disaster response scenario (Fischer et al., 2014). The design was similar to Moran et al., with the addition of moving “radiation clouds” visible only to an online coordinator. This added spatial and time pressure to the process of “rescuing” items and people and moving them to drop-off points. While this design still does not manipulate variables, it's easy to see how changes in communication tools, or in the degree of pressure provided by the “clouds,” could have been implemented. The use of iterative prototyping prior to the game launch is also notable, and may be a useful model for those considering designing such studies in the future.

What Can We Learn From Other Applications of ARGs?

Research is far from the first field to take advantage of the ecological validity and motivational power of ARGs. Indeed, the prototypical modern ARGs were created with a very specific goal in mind: outreach and marketing for movies and video games. Perhaps because ARGs have not tended to be profitable in their own right, large ARGs in particular are often designed for something beyond pure entertainment. ARGs have been used to plan and explore foresight scenarios, raise funds for causes, direct volunteers, and motivate people who want to exercise more. Supported by an increasing body of successful case studies, ARGs have become widely used for education and training. By looking at these applications, we can draw important lessons for expanding the use of ARGs to a new domain.

Marketing

While there are a number of precursors and intermediaries, the earliest commonly recognized “classic” ARGs are *The Beast* and *I Love Bees*. Both were created as a means of engaging players more deeply with other entertainment products: the *A.I.* movie in the first case, and the *Halo 2* video game in the second. These games drew in players through rabbit holes in both product-associated and seemingly independent settings. *I Love Bees*, for example, included links to the titular website in “subliminal messages” in the video game trailer (rewarding close frame-by-frame examination) and on jars of honey sent to players of previous ARGs. Both were considered extremely successful, with *I Love Bees* in particular credited with contributing to *Halo 2*’s record-setting first sales day.

While marketing ARGs have continued to focus primarily on supporting more traditional entertainment forms, they have also been used for products ranging from t-shirts (*Edoc Laundry*) to cars (Audi's *Art of the H3ist*). Disney's creation of *The Optimist*, and plans for further games under the Living Worlds banner (Disney, n.d.), provide evidence that this type of effort is appealing and effective for even the most successful of brands.

Engagement is at the center of marketing ARGs. As described above, narrative is psychologically powerful, creating meaning and identity around connections that might otherwise appear unimportant. Participation in a narrative game connects people's self-concepts to other aspects of that game, as well as to the process of play itself. Play can lead to feelings of commitment, and a willingness to invest effort and action—something that other types of applied ARGs take advantage of as well. For a research ARG, this type of psychological power can be directed both towards maximizing retention over the course of an experiment, and encouraging participants to respond to in-game situations in an ecologically valid manner.

Scenario Planning and Exploration

ARGs are a unique way to embed people in counterfactual and hypothetical scenarios, offering an added dimension to foresight and planning exercises. Putting these exercises in a game context can increase motivation and/or give participants a deeper and more detailed perspective on possible futures.

Scenario planning ARGs come in two basic forms, both pioneered by the Institute for the Future (IFF) think tank. The first, relatively freeform version, is best represented by *World Without Oil*. Atypically for an ARG, *World Without Oil* involved no mysteries or puzzles. Players were given a basic peak oil scenario, and asked to fill in the details—from their own point of view as someone living through the first weeks of a long-term fuel shortage. They did this through in-character blog posts, videos,

e-mails, and other online artifacts produced within the scenario world. The end result was a crowdsourced corpus of detailed suggestions and stories about how individuals, families, communities, and governments might respond to such a crisis. This provided the game's sponsors with a better understanding of second and third order problems that could arise from the shortages, and of potential solutions shaped by local needs and resources.

The second, more constrained form is IFTF's Foresight Engine model, a gamified hybrid of microblogging (e.g., Twitter) and more traditional ideation exercises (Institute for the Future, n.d.). This form is less immersive, but more organized and still an effective method of motivating participation and creating smaller narratives. Over the course of a few days, participants play virtual "cards" on which they briefly describe possible developments in a specific field (e.g., in *Maker Cities*, exploring how current trends in the DIY community could shape urban futures). Other players then add cards that build on this foundation with critiques, elaborations, and potential roadblocks and solutions (For an example see images in Tester, 2013). Each contribution gains points and recognition as players argue for their predicted futures. The Navy's adaptation of the Foresight Engine for their *MMOWGLI* games (Hagan, 2012), to explore scenarios from combating piracy to leveraging reserve troops, suggests that organizations of many types see significant potential in this model.

Both of these types of games demonstrate the ways in which a good narrative, and an effective set of game mechanics, can make counterfactual scenarios come alive for participants. *World Without Oil* bloggers treated their imaginary oil shortage in a serious and psychologically valid way, and filled in details that organizers had not originally considered. Rich interactions can develop from the mix of cooperative and competitive behavior: participants both work together to create visions of the future, and compete to advocate for opposing versions of the story. This immersive potential suggests some ways that research ARGs might increase the

plausibility and realism of the interactions they are intended to study.

Attitude and Behavior Change

As with many other marketing techniques, designers have appropriated ARGs for use in “selling” causes, healthy behaviors, and other changes intended to benefit individuals or society rather than a particular brand’s bottom line. Many have been successful, but there have also been cases of backlash when target communities perceived a game as manipulative or deceptive.

Conspiracy for Good set a notable example of both the potential and pitfalls of this type of game. Combining on-line puzzle solving with interactive street theater, it set players against an evil international corporation that had stolen funds intended for the creation of a library in Africa. Fundraising through the game was used to build and stock five rural Zambian libraries, as well as donate books elsewhere and create 50 college scholarships. With hundreds of thousands of online players and hundreds of live players, the game was a striking success.

However, there was little overlap between the two groups: the international population with time to follow the story and solve puzzles over several weeks was not the same as those able to attend multi-hour live events in London (Stenros et al., 2011). Live players often arrived with little familiarity with the story, and expectations based on non-narrative street games (e.g., the Come Out and Play festivals) rather than ARGs. This suggests that ARGs, if they don’t develop broad outreach and recruitment methods, may attract their own relatively narrow population samples, shaped by the type of activity offered and their perceived genre and format. This could be an advantage for studies focused on the interested population (e.g., a game built around team formation for crisis management might attract participants who are particularly motivated to act in crisis situations), but studies looking for a less self-selected population may wish to tailor their recruitment tactics accordingly.

Conspiracy for Good also demonstrates the dangers of the TINAG (“This is Not a Game”) aesthetic. Rabbit holes for the game included a “leaked” video from evil megacorporation Blackwell Briggs, which attracted the attention of real-world activist groups that oppose wrongdoing by large corporations. When Nokia’s involvement with the design became public, these organizations interpreted the game as a hoax intended to distract them from real-world infractions (e.g., Nokia’s own provision of surveillance technology to Iran), and took it as an invitation to disrupt Nokia’s game-related websites (Cutforth-Young, 2013). These actions illustrate the sort of backlash that can occur when game-related information goes unmarked, or when potentially controversial organizations are not open about their sponsorship.

Understandably, games aimed at individual rather than community change may often have a very different format. These games can allow multiple people to follow the story at their own pace and solve problems in parallel, rather than asking them to work in concert. For example, the *Zombies Run* mobile app provides players with a dramatic narrative, during which they must periodically solve problems—and run to escape zombies. Run times and distances increase to keep pace with players’ fitness levels. The game has been extremely successful—downloaded and used by upwards of 750,000 people. It has recently been joined by a sister app, *The Walk*, a spy thriller focused on movement at a more sedate pace.

Coral Cross represents a hybrid of behavior-oriented and foresight games. Originally planned by the CDC as a pandemic preparedness exercise in the style of *World Without Oil*, it was upstaged by current events when H1N1 hit the news just prior to the expected launch date. Designers quickly retooled it with a focus on education and preparedness for existing illnesses, helping people to respond in healthier and more effective ways. This type of flexibility may be necessary for any game working with issues that are, or may become, immediate concerns in the real world, and that may include some research ARGs as well.

Education and Training

Educational ARGs have moved well beyond the stage of the first notable pilot projects, and into regular practice. Examples include:

- *Reality Ends Here*, used to orient new media students at the University of California and encourage them to form creative collaborations.
- *Ghosts of a Chance*, providing a game-based framework for exploring and understanding the Smithsonian American Art Museum collection—like the fitness games described above, played by individuals in parallel.
- *Arcane Gallery of Gadgetry*, in which students communicated with the past to learn about the history of the US Patent Office.
- *Tower of Babel*, a multi-school European game intended to increase motivation for learning foreign languages.
- *RevQuest*, Colonial Williamsburg's game that teaches visitors history while guiding them to experience the site and interact with historical characters.

Connolly, Stansfield, & Hainey (2011) list several features that give ARGs educational value, many of which generalize to other applications:

- Offering problem solving opportunities at a range of levels lets students select their own starting points. For a research ARG, this strategy could be used to study the same type of psychological process at different levels of skill and experience.
- Progress and rewards (e.g., leaderboard placement), common in games regardless of goal, can be used to support student assessment--and research metrics and measurement.
- Narrative devices such as character and plot fit well with subjects that draw on history, literature, or current events. Just as they can increase player understanding of these topics, they can also be used as a framework for studying related social dynamics.

- An active community can increase engagement and drive recruitment of new participants.

However, these advantages can easily be lost through poor design or outreach. Moseley, Whitton, Culver, & Piatt (2009) document several less-than-effective educational games. These games failed for a number of reasons. The TINAG aesthetic proved to be a pitfall again, this time because students unfamiliar with ARGs did not recognize from publicity materials that they were being invited to play a game, understand that it was meant to benefit their studies, or know what sort of rewards it might offer for participation. In another ARG, a poorly calibrated first challenge proved too difficult and led to many drop-outs. Poor challenge access, lack of alternate paths or hints for “stuck” players, and failure to convince players that game organizers were using fair evaluation techniques, were other factors that prevented successful engagement and retention.

ARG-like activities have also been used for training in professional and pseudo-professional settings. The Goruck backpack company runs “challenges” in which teams must complete arduous physical tasks, or simulate intelligence operations (Goruck, 2014). Military war games, used for training and planning, often have ARG-like aspects. ARGs have also been used for corporate communications training (Moseley, Alexander, & Bono, 2010).

Given their focus on increasing educational participation and outcomes, the common lack of rigorous evaluation metrics for these kinds of ARGs appears somewhat surprising. While there is considerable anecdotal support for improved motivation, engagement, and subject matter comprehension, there have been few rigorous comparisons between ARGs and other novel, interactive teaching methods. It seems likely that creators of research ARGs could benefit from collaboration with educational ARG creators on metric development, and vice versa.

What Do Research ARGs Need For Success?

Where Are We Now?

In the years since *The Beast* and *I Love Bees*, ARGs have become more diverse in who creates them, intended scale, and player base. The palette of design choices has broadened as organizers experiment and develop new techniques. Organizers have experimented both with new technologies (e.g., augmented reality) and low-tech games (e.g., for outreach by aid workers).

For the first modern ARGs, corporations set aside 7-figure budgets to draw in large international player populations. These expensive, large-scale games still take place, but are by necessity relatively rare, and generally associated with major advertising campaigns. *Art of the Heist* holds the record with a \$4 million price tag. Cheaper and smaller-scale ARGs are, however, becoming increasingly common. The educational ARGs described above are one factor pushing this growth—an ARG for a single history classroom might cost only \$100 to build, discounting the cost of the educator’s time (Borden, 2012). Another factor is the increasing number of “grassroots” games, which may cost between a few hundred and a couple of thousand dollars (IGDA, n.d.). These are often run entirely online, by an individual or small group who simply want to create an interesting experience.

Research ARGs are likely to fall somewhere in the middle of this range, with the bulk of funds going toward staffing. Interdisciplinary collaboration will be needed, with game designers working with social scientists

to create playable, informative experiences. Most designs are likely to also need a sizable support staff; previous ARGs of comparable scale have required between 5 and 50 staff. Both staff and budget size have a major influence on how quickly new material can be released in response to player actions. Even if the general outline of the ARG is known in advance, player actions are still likely to shape the game throughout—meaning that better staffing allows responses to be tailored more quickly, and with greater psychological validity (L. Fischer, personal communication, October 22, 2013).

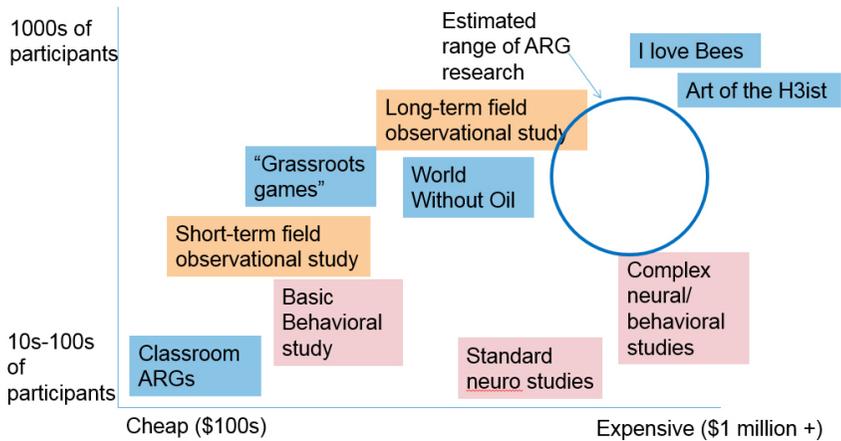


Figure 1. Cost and Scale Comparisons for ARGs and Research Activities.

ARGs vary widely in the degree to which they record demographic information about players. The largest and most successful are the most likely to do so—and these, at least, appear to gather a reasonably diverse audience (Dena, 2008a). Games that aren't specifically marketed towards a male population tend towards a 50/50 gender split. Players come from a range of countries, although the trend is towards English-speaking nations (as most ARGs described here are produced in English). For

example, 91% of *Perplex City* players came from the US and Canada. The independent game Metacortechs logged participants from 115 countries, but did not list the percentage from each, or how many locations were responsible for the majority of players.

The Lost Experience is notable for drawing extensive player populations from North and South America, Europe, and Asia. To get this level of diversity, their team included 8 multilingual puppetmasters from around the world (McGonigal, 2008). In creating this community, they found that players from different countries played with very different styles—for example, players in Spain and South America sought more full-immersion role-playing experiences, even when discussing puzzles with other players, while US players favored a more direct approach to problem solving in which they could discuss the logic of the story from an outside perspective. Other designers have found similar cultural differences. American audiences tend to be most interested in competition, while live events are especially popular in Europe (Phillips, 2012).

ARG design trends have changed over time, becoming less dogmatic and more diverse. This is illustrated by a change in tone and scope of the descriptive literature. In 2005, the seminal work was Dave Szulborski's *This Is Not a Game*, reflecting the TINAG aesthetic that was considered central at that time. A similar aesthetic is reflected in the International Game Design Association's 2006 white paper. In 2009, *Pervasive Games: Theory and Design* (Montola, Stenros, & Waern) reflected a broadened view of the field, and an increased emphasis on the scope of settings in which games were played, rather than the precise way in which they were presented to potential players. Just as movies need not present themselves as documentaries or found objects (e.g., *The Blair Witch Project*) in order to have audiences on the edge of their seats, ARGs can carry an overt fictional label and still effectively immerse players in their story worlds. *A Creator's Guide to Transmedia Storytelling* (Phillips, 2012) follows this trend, explicitly discouraging TINAG techniques.

Many early ARGs emphasized puzzle-solving, with narrative providing a framework and justification for that activity. Newer ARGs may still involve puzzle-solving, but may also involve storytelling, coding, filling out surveys, and complex social interactions in which NPCs or other players must be persuaded to take certain actions—in other words, the story now provides a framework for a much wider range of activities and challenges that players may find enjoyable and engaging in the right context.

Transmedia presentation remains a central aspect of ARG play, with most games requiring use of a diverse set of technologies and venues. Live play continues to be a powerful feature that can add intensity and verisimilitude to play—although, as currently designed, only a small percentage of participants are likely to attend. At the same time, common in-game technologies have changed to match availability—current games are far more likely to require smart phones than pay phones.

This trend is likely to continue as new technologies become available, and many emerging technologies could provide interesting opportunities for ARG designers:

- *Wearable computers.* Technologies such as Google Glass and smart-watches will make it easy to share different pieces of information with different players, even in the same space. This allows even small-scale games to include challenges in which players must collaborate to share information, with real-time updates changing what is available to each player.
- *Augmented reality.* Enhanced reality tools will allow rooms to contain virtual objects, or information to be “hidden” in everyday locations. Google Ingress is an early example of this, with public artwork enhanced to create the “portals” central to game play.
- *Internet-connected objects.* The “internet of things” could complement enhanced reality, allowing physical environments to change in response to the presence or actions of players.

- *Portable sensors.* New sensors, independent or connected to smartphones, could be used to give players artificially enhanced senses, to allow puppetmasters and environments to respond to player emotions (e.g., doors that open when you calm your heart rate), or to record valuable neurophysiological data for research.
- *Maker spaces.* The increased public availability of 3D printers, modifiable electronics, and other high-tech DIY tools also increases the number of problem solving strategies available to players. Future challenges, for example, could require players to reconstruct and print out a carefully shaped key to retrieve a clue.

One thing all these technologies have in common is physicality: they are new ways of making technology pervasive rather than limiting it to particular spaces, times, and devices. This suggests that future ARGs will return to the tradition of geocaching, and like Ingress and Zombies Run, take increased advantage of the physical world rather than relying primarily on online interaction. Live play, rather than a difficult-to-access major event that must be carefully staged by game-runners, may instead become something that most or all players are able to do in the places and times available to them. This has the potential to become a major advantage for ARG-based research, and could expand our ability to study face-to-face social interaction in greater depth. It may also allow for extensive comparison between populations in multiple locations.

Best Practices for ARG-Based BSSR Research

We have now explored the common characteristics of ARGs and related activities, and discussed how these play out in a variety of settings and with a range of applications. As described, which characteristics of ARG design may matter most for research? And which limitations of current lab-based research can they help overcome?

BSSR Research Issues	Lab Research	Field Research	ARG-Based Research
Large-scale, complex social interactions	Interactions usually small-scale & simplified	Full-scale and in context	Large-scale and complex, but limited to game setting
Participant pool size & diversity	Frequent focus on WEIRD participants, often small Ns	Limited only by scope of study	Potential to draw large #s of participants from around the world, limited by tech access
Ecological validity of setting and behavior	Multiple published critiques	Generally viewed as high	Potentially significantly higher than lab, and close to field under some circumstances
Data collection and analysis	Potential for thorough and rigorous data collection & analysis	Variable, but usually more limited than lab – most analyses are descriptive and correlational	Some logistical and ethical limitations, but potential for extensive high-quality data collection & use of modern ‘big data’ analytics
Experimental control	Potential for high-quality control of IVs and control variables	Rare and minimal	Quasi-experimental designs may be needed—but control potentially easier than in field

Table 1. Comparison of ARG-based research potential with traditional methodologies

Large-scale, complex social interaction is a major area in which ARGs could provide new opportunities for research. These interactions can be notably difficult to scale down meaningfully in a lab setting, as well as to get a complete picture of in the field. A well-designed ARG can create a community from the ground up, allowing dynamics of cooperation and competition, faction formation, information spread, culture formation, and problem solving to be observed from their inception, under somewhat controlled circumstances that can help to make otherwise unwieldy data much more interpretable.

ARGs also have the potential to recruit and retain larger numbers of participants, with greater diversity, than other controlled studies. Depending on the set-up, recruitment may be viral, or word-of-mouth may be supplemented by more organized outreach. While a research ARG may be unlikely to draw the numbers of *I Love Bees*, few experiments can boast even the several hundred participants gathered by many mid-sized ARGs. Because an ARG is entertainment, even when it has other overt goals, its rewards are of interest to a larger population than a lab study offering course credit. Because an ARG can be played in convenient locations, it is accessible to a larger population than lab-focused studies as well. Because it offers an entertaining experience, it may also permit that population to be recruited at a lower cost per participant. However, researchers may want to consider carefully what types of participants are likely to self-select for a particular game, and whether they provide an appropriate sample for the study questions—and recruit accordingly.

Finally, ARGs can ask players to interact with the world using familiar tools, and in familiar places. Even when the purpose of these tools is unusual, behavior may still be more ecologically valid than in the artificial environment of the lab. Examining a website for signs of alien invasion may be unusual—but probably not as unusual from a behavioral and psychological standpoint as engaging in a forced-choice visual perception task while wearing an EEG monitor. Just as the CDC (2012) uses a zombie apocalypse to stand in for many more probable scenarios in disaster preparedness activities, that alien invasion could stand in operationally for phishing attempts, provision of inaccurate information, sponsorship by a particular organization, etc.

All of these characteristics provide opportunities for high-quality research ARGs, but also carry pitfalls. Both ARGs and research studies can be complex endeavors, and the combination promises both new risks and new potential. Experience with both forms suggests some specific best practices

that could improve the likelihood of a successful hybrid project.

Best Practices for Design. The degree to which ARG-based studies can approach the rigor and control of lab-based research remains an open question. The potential for manipulation of independent variables, control of confounds, and randomized assignment to conditions may vary depending on the hypotheses, populations, and design constraints specific to a given research project.

Manipulating independent variables may be the easiest of these challenges to address. Many standard aspects of ARG design could be varied systematically for research purposes. In some cases this could look much like the A/B testing often used by outreach professionals: between-participant variations in website appearance or content in order to measure changes in impact (Christian, 2012). Other cases might be more involved: different participants might be offered access to different tools, receive different incentives for specific strategic choices, or encounter characters who frame a conflict in different ways. Longitudinal manipulations could also be possible: interventions that attempt to influence team formation, for example, or changes in available resources over the course of the narrative.

Randomized assignment to conditions may also be possible for many designs. *Ingress* players currently choose a faction upon joining the game; however, early faction assignments were random. Similarly, participants in ARG-based research could be randomly assigned to teams, instruction sets, or other starting conditions. In some cases, depending on data collection protocols, it may be possible to make these variances blind for both observers and participants. In others, this flexibility may be limited.

The latitude that ARG players have in their possible strategies may limit randomization—for example, it might be difficult to entirely prevent participants from switching teams following initial assignment. Variables that

intersect with non-game choices and events (e.g., location, interactions with other players) would also be difficult to randomize.

Control for potential confounding variables is often a significant challenge for field studies, and is likely to be so in field-like ARG studies as well. With participation embedded in everyday life, many variables simply fall outside the realm of experimenter control. However, it may at least be possible to account for specific variables that have been identified as concerns. Recruiters could seek out gender balance with and between teams, for example, and players could be given experimenter-owned devices to avoid effects of using different operating systems or software tools.

Given the potential for at least some randomization and control for confounds, not all ARG-based research may fall under the strict definition of quasi-experimental design. Nevertheless, lessons from quasi-experimental research may help to improve study designs, maximize control where it is possible, and account for its imperfections when developing data collection and analysis strategies.

Best practices for increasing motivation and engagement. Motivation, engagement, and retention, are challenges for any BSSR study lasting longer than a single session (and often for a single session depending on the task at hand). They are also challenges for ARGs—but successful games have been shown to perform far better on this front than lab-based research, because a positive player experience is central to their goals. While research may have to prioritize the collection of valid data, a successful research ARG should consider ways to combine engaging game play with practical outcomes. This combination of play with practicality (although not with rigorous data collection) has been dealt with effectively in many of the applied games described above, particularly educational games.

Researchers have identified a number of motivational factors for such games (Moseley, 2008; Moseley et al., 2009; Whitton, 2009):

- Regular opportunities to complete challenges appropriate to players' skill levels, and feedback and rewards for progress through the game.
- Engaging narrative and an opportunity to learn more about the story through play.
- Opportunities for competition with other players.
- Opportunities for community-building and cooperation.
- Players feel that they have influence over game outcomes. Note that perceptions do not always match the reality of the game: many designers discuss the needed balance between player agency and the creation of a dramatic and effective narrative. The “string of pearls” structure, in which episodes are set in advance, but players can influence the course of a given episode, is one possible compromise that has been successfully adopted in previous ARGs (Garrand, 1997).
- Opportunities for creativity, either through problem solving or the development of game-related artifacts.

A given game need not have great strength in every single one of these areas—for example, games such as *World Without Oil* emphasize narrative and cooperation over puzzle-solving and competition.

Ordinary games depend on extensive play-testing in order to produce a successful combination of motivational factors, and iterate play with small test groups until they achieve their desired experience. This is the equivalent of pilot testing in research. Due to their length and complexity, ARGs are notoriously difficult to playtest. Designers often resort to on-the-fly modification response to feedback during play. Research design requirements—and the requirements of IRBS—are likely to increase the difficulty of this type of modification. Research ARG teams may need to develop new ways of testing their proposed designs with small groups, rather than relying on the flexibility available to other ARGs.

The “String of Pearls” Structure

The String of Pearls structure is a common strategy for advance planning of stories that will be, or will appear to be, influenced by player input. It also has important implications for experimental design.

Designers may plan an interactive story with periods of free play interspersed with “episodes.” Each episode has a known problem for players to solve, set of possible strategies, and a known end state that will result from any plausible solution. This allows designers to set up multiple episodes in advance while still giving players room to make interesting and meaningful choices. The structure can vary, for example, by having a finite number of possible endings, with branching possibilities stemming from those different endings (e.g., following episodes play out one way if players choose to fight an adversary, another if they make peace).

Research designers can take advantage of this structure by using episodes to expose players to specific independent variables, or to take measurements during specific activities.

One option is to take the iterative playtesting model, and adapt it to specifically address research needs. For example, a series of short ARGs might be used to first produce hypotheses, and then test them, with each ARG (or episode within a larger ARG) producing data that would be used to inform the next one. Such a strategy would allow in-depth exploration of BSSR phenomena to take place in parallel with improving the play experience based on feedback.

Best practices for increasing participant diversity. As described above, ARGs draw a somewhat diverse population, but most produce a sample heavily biased towards the US and Canada or the UK—likely because they are created and run by people from these countries, in English. ARGs that

make the effort to take other cultural perspectives into account, and/or present their material in other languages, draw a more complex crowd.

Regardless of this variation, ARG players are likely to be high socioeconomic status, with significant leisure time, freedom of movement, and access to technology. In other words, they are likely to be WEIRD, but perhaps less so than the university-attending “standard research subject”. Further, ARGs provide opportunities to push these boundaries, and to reach at least some populations that don’t normally participate in BSSR research.

Perhaps the most important indication of the potential for increased diversity is the experience level of players. A sizable community of ARG enthusiasts actively seeks new games to play, scouring media for rabbit holes. Nevertheless, most *Perplex City* and *The Lost Experience* players were new to ARGs. It is quite possible for good outreach and creative use of communication venues to attract a wide range of players.

When planning outreach, it’s also necessary to remember the tiers of player participation discussed above (Dena, 2008b; Phillips, 2012). Active players, who can provide data, are a small percentage of those who will simply follow along passively to whatever degree is permitted. Interaction can be increased by minimizing barriers to participation, but as long as play remains a volunteer activity, a great deal of self-selection should be expected.

Best Practices for Storytelling. The role of story in a research ARG is potentially ambivalent. In the scientific community, narrative is often seen as a framework to get people interested in information that might not otherwise engage them. It would be easy to assume that its major purpose would be to “dress up” research participation and make it more appealing. Indeed, people tend to find tasks and information more interesting when they are given a storytelling framework (Mott et al., 1999), and

such a framework is likely to be effective in engaging participants. However, narrative may not best be seen as an interchangeable “skin” laid on top of research activities. Narrative engages people because it is deeply embedded in our psychologies—and the precise narrative used shapes the way we think about information or participate in an activity. In other words, narrative choices will be an integral variable influencing the interactions being studied. Different narratives may prime the use of different social “scripts,” and thence different behavior (Brouillette, 2009).

Research in this area is longstanding enough to maintain confidence that the effects of narrative on behavior are real and sometimes dramatic—but findings are mixed enough to create uncertainty about the precise effects of that influence. This is an area of interest to many communities (e.g., the DARPA Narrative Networks program) and may, in fact, be an important area for early ARG-based research to explore. Pending these more specific findings, potential researchers are urged to treat narratives as important variables, and design them with the care that would go into creating a new piece of experimental equipment.

Another aspect of storytelling that will be important to ARG-based research is character. ARGs often depend on a few puppetmaster-directed characters to personalize their stories, whose villainy or peril motivates players to deeper emotional and cognitive immersion. For example, *Conspiracy For Good* pits the activist Nadirah against nefarious Blackwell Briggs CEO Ian Briggs. Both hero and villain were played by game organizers at in-person events, by video chat, and over e-mail. In game terms these are *non-player characters (NPCs)*, and represent both a tremendous investment of time and effort, and a central part of what makes play appealing. The use of NPCs is not far from researchers’ use of assistants who follow a script while pretending to partner with real subjects—for example, the “victim” in the Milgram (1963) shock experiments.

Researchers in recent years have often designed studies in which such an

actor can be replaced by a computer—either following a set script, or taking a non-verbal role in a game where each player must make specific choices. Although intelligent agents have grown more sophisticated, this is more problematic for an ARG, in which characters must be able to respond to a wide variety of possible player actions in different media. The complexity of social interaction that distinguishes ARG-based research from lab research also minimizes the degree to which interactions of interest can be automatized. The human brain is still the best platform for modeling and simulating many types of human interaction.

Best Practices in Staffing. An ARG-based research team is likely to draw on an array of experts, with specialists in multiple fields all playing necessary roles in successful coordination. Some have suggested that the research team of the future will look a bit like a movie crew, and the disciplines needed for ARG-based research certainly reflect this prediction.

- *Research Experts.* Needless to say, the team will require experts in whatever areas of social, behavioral, and neurological science inform the research topics and measurements. However, ARG-based research, regardless of topic, will also benefit from the presence of anthropologists or ethnographers familiar with embedded observation techniques. Even with more quantitative metrics in place, these techniques are a necessary component of interpreting player behavior and responding to it appropriately. Trained research assistants to assist with game-running, and take on hypothesis-relevant NPC roles, will also be a plus.

The data coming out of research ARGs are likely to be complex and multi-faceted, and dedicated statisticians and analysts are likely to be needed. On-site ethicists, familiar with the guidelines for human subjects research, may also contribute to appropriate choices in an environment where design changes must be fast and responsive.

- *Game Design and Implementation Experts.* ARG design is a specialized art form, and research ARGs will need experienced game designers and puppetmasters to complement the skills of the researchers. These experts can mold a good research design into an ARG that is engaging, runs smoothly, and avoids major pitfalls.

ARG-based research teams may also wish to make use of other artistic specialties. For example, documentary makers can film, and otherwise gather and organize data on, in-game conversations and interactions. Outreach and publicity specialists can bring the game to the attention of populations of interest, respond to concerns, and facilitate positive coverage.

- *Citizen Scientists and the Crowd.* The amount and complexity of data likely to be produced by research ARGs provides an intriguing opportunity: to supplement team statisticians with crowdsourced analysis. There are several ways this could be managed:
 - As described above, many people who follow ARGs prefer passive observation to active participation in the story. However, these lurkers may be willing to participate at a lower level. One possibility is to ask them to perform simple analysis of player-produced interactions. This type of analysis can be gamified, as it has been on Zooniverse, where players categorize galaxies or wildlife. In cases where human sensory judgment outperforms automated analysis, this can be extremely effective.
 - Active ARG participants could analyze other participants, either as part of play (e.g., providing a report on a rival faction) or afterwards.
 - Game-produced data could be made available to a post-game analytics challenge. Challenge contests offer prizes to members of the public who are best able to address a prob-

lem—in this case, producing the most informative analysis of a large, complex data set. This type of analysis activity is particularly effective for gathering multidisciplinary input and experimenting with a wide range of potential analytic strategies.

ARG-Based Data Collection

Research ARGs provide a number of opportunities for improved data collection. Increased complexity over lab research combines with the increased controllability over field research to produce novel advantages.

ARGs make it possible to collect detailed information about a wide set of variables. With player permission during informed consent, and by focusing on key venues where interaction takes place, researchers can record and analyze information about conversational content, tone of voice, body and language in near-real-world environments. Montola, Stenros, & Waern (2009) discuss some of these possibilities, and their limitations:

- Cameras and microphones can record player activity—however, this becomes more difficult if players are moving around, and often produces poor-quality records of multi-person social interaction. There are also legal limitations on recording in public spaces, even with player consent.
- Dedicated devices or programs for player use can record log files on technology-mediated communications and actions.
- Player self-reports can be requested either as part of the game or as a separate activity. The Babylon system (Waern, Ahmet, & Sundstrom, 2009), for example, provides a graphic interface where players can log both their actions and emotional responses. While it improves the quantity and depth of data, participants also report some disruption to their play experience. Reporting within the context of the story is less disruptive to play, but carries its own risks—for example, players will tend to “narrativize” their reports of events, telling an entertaining story

rather than giving a careful and accurate summary.

- Online player forums are a long-term record of not only specific social interactions, but players' plans and interpretations of game materials. These interactions are often on a 'meta' level, with players speculating about, and trying to second guess, puppetmaster intentions.

Informants and game-runners can also embed themselves in player groups, providing detailed information about player interactions. Stenros, Waern, and Montola (2012) discuss this possibility in greater detail. They emphasize that while not every recording and analysis method must involve this kind of participation, participatory methods are a vital component of any analytic plan. Including researchers in the process of play ensures that records will be understood in the context of player experiences and player-created culture. For example, embedded researchers may be able to pick up on subtle cues as players shift in and out of the game's in-story framework while discussing strategy. As described above, the pattern of these shifts may vary between games, or between players of different backgrounds.

Analysts working with the *I Love Bees* player forum engaged in extensive content coding, analyzing language use within each user-created discussion thread (Gurzick et al., 2011). This type of analysis draws on standard discourse analytics methods in which researchers categorize semantic and other content, but extends these methods to understand exchanges in the context of game strategies and the emergence of the ARG's storyline. Gurzick et al. focused in particular on the emergence of collaborative problem solving, using text analysis to explore the dynamics as players created, organized, and dissolved spontaneous teams in response to in-game needs.

Network analysis is another method well-suited to ARG-produced data. The organizers of *Reality Ends Here* studied the formation of creative teams, and connections across teams, over the course of the game (Stokes,

Watson, Fullerton, & Wiscombe, 2013). This game was deliberately designed to foster collaborative creation of multimedia art, and is a good example of how ARGs can be used to both produce and study complex, ecologically valid social interactions. Results supported and extended previous findings in network analysis—students with greater network centrality, who acted as creative hubs, also had higher game scores (reflecting greater productivity and creativity).

A key advantage of ARGs in this arena is their multi-modal presentation and interaction across media and settings. These cross-modal communications and relationships can provide the type of multidirectional input that is characteristic of real world events and people. However, *Reality Ends Here* illustrates the advantage of central reporting for these complex interactions—players used the full range of tools available to them for their collaborations, but ultimately brought their artwork to a central office for scoring.

ARG data collection faces 3 key challenges:

1. Some types of data collection can be intrusive, interfering with the ecological validity of the scenario. Self-reporting, if it takes place outside of the story, may distract from, and thereby change, in-game interactions. Players' awareness of surveillance may also affect interactions and lead them to deliberately "perform for the camera," while a lack of awareness has serious legal repercussions. Players may try to identify or fool embedded researchers as part of their strategy.
2. As described above, players reporting on their own experiences may attempt to revise them into game-appropriate narratives, increasing reported drama or leaving out details that don't seem to fit. They may also treat manipulation of in-story data collection as part of game play—for example, providing false answers to a survey in order to try and reach an in-game goal.

3. Collection during complex interactions can be logistically, and sometimes legally and ethically, difficult. Face to face interactions are of particular interest for ecological validity, but may be difficult to record if they don't take place under carefully controlled circumstances. Cross-modal discussions may be difficult to follow, and key components may be missed as players seek new tools and venues for reaching their goals.

Even during non-research games, puppetmasters encounter these issues. Experienced game-runners report a variety of compromises. Carefully discrete embedded observation is generally agreed to be necessary. Even participants who have signed informed consent agreeing to surveillance may not pay it much attention during actual play, particularly if the story is diverting enough. Montola, Stenros, and Waern (2009) recommend having a story-based explanation available if these things are discovered during play sessions. (For example, a hidden camera might include a clue pointing to the identity of adversaries.)

Ultimately, while all methods of recording data during ARG play have their limitations, a well-chosen selection of complementary strategies can provide parallax.

Ethics of ARG-Based Research

The ethics of human subjects research are a well-explored area—albeit still one that produces ongoing controversies and debates. The standard aspects of these requirements (informed consent, justice, benevolence, etc.) do not have to be repeated here, but should be assumed to apply to ARG-based research just as they do to lab and field research. However, the use of pervasive games may raise new questions in each of these areas, and may add a few ethical challenges of its own.

Pervasive game designers have explored many of these issues on their own terms. For example, informed consent is a major issue when players

are to be recorded—but knowing about some types of surveillance may interfere with entertainment goals (Montola, Stenros, & Waern, 2009). For example, surreptitious monitoring of heart rate might allow the creation of a door that opens when you meditate in front of it. More subtly, as described above, knowing that one is being watched may change one's behavior. While game designers sometimes err on the side of immersion, researchers would reasonably be expected to follow the stricter and more legally constrained guidelines of their own field.

Informed Consent. The TINAG aesthetic is almost diametrically opposed to informed consent—early ARGs pulled players in organically, never openly acknowledging the type of activity in which they were engaged. Fortunately this is neither universal nor necessary for a compelling game experience. Many modern games more explicitly mark game entry, require direct agreement for participation, and/or provide non-intrusive markers to distinguish game websites from ones that originate elsewhere. Some Nordic live action games require player contracts that protect the rights of all involved, while allowing play to proceed easily following agreement (See for example Stenros, 2008).

Game designers sometimes worry that agreeing to specific activities undermines the use of exciting surprises as storytelling devices. Those interested in ARG-based research have also expressed concern that overly specific informed consent could undermine research goals. Fortunately, social science research has dealt with this issue as well—not out of concern for entertainment, but from awareness that many psychological measures can be profoundly influenced by participants' knowledge about what they are doing. A lack of specific detail in a consent form or debriefing, sometimes even shading into deception, may be considered acceptable if 1) no harm is done by the lack of information, and 2) people would reasonably be expected to agree to the activity if they did know more details. For example, older adults often perform worse on a test if it is described as a memory exam (Hess, Auman, Colcombe, & Rahhal, 2003). A consent form might therefore reasonably state that participants

will be asked to complete “general tests of mental ability,” even if the specific purpose of those tests is actually to examine memory.

Experienced pervasive game designers follow this same logic, suggesting that consent forms focus on the general types of activities that may be involved (Montola, Stenros, & Waern, 2009). Players might be told that a game requires strenuous physical activity, but not that it involves climbing a wall to escape from zombies at a climactic moment. This does require that opportunities for renegotiation be accessible and known to players throughout the event. This could be as simple as giving players a hotline number to reach puppetmasters—a player who discovered that a particular activity was problematic could call and discuss options, or if necessary withdraw their participation.

One aspect of consent, highly relevant to ARG-based research, is open discussion of a game’s sponsorship and purpose. As demonstrated by the negative reaction to Nokia during *Conspiracy for Good*, many players do care who produces an ARG. This may be because they prefer to affiliate with organizations whose actions they approve of, because they trust some organizations more than others to run games ethically, or because they don’t want to invest their time with a poor storyteller. Similarly, many players are likely to care about non-entertainment purposes of a game, and may prefer to choose whether they will invest their time or offer up their data in support of a given goal (Cutforth-Young, 2013). In spite of the potential advantages of subjects who are unaware of a game’s research purposes, experience suggests that these advantages are unlikely to meet criteria for permissible deception by either an IRB or the ARG players themselves. ARG researchers may wish to communicate early and clearly, when potential players give consent, about their work’s sponsorship and purpose.

One advantage of using informed consent as a gate to play is that it more easily facilitates out-of-game discussion of the findings. When IARPA’s

UAREHERE RFI went out to ARG designers, at least one initially wondered whether it might be a (relatively dull) rabbit hole in its own right (Cutforth-Young, 2013). If game material is known to be clearly marked, discussion of study design and results can proceed more freely without participants worrying about whether they should provide strategic responses for the sake of in-game goals.

Privacy and Data Collection. If ARG-based research is to take full advantage of its potential, it will very likely involve making records of communications and interactions in a variety of media, as well as in face-to-face meetings. At minimum, researchers should expect that this will require clear informed consent. In addition, there could be places in ARGs where it may be appropriate to reduce or eliminate data collection for the sake of preserving player privacy.

One option for safeguarding privacy could be to require players to engage in game-related communications solely through dedicated accounts or devices. This approach should clearly delineate the boundaries within which players may expect to be observed, although it also opens up the possibility of players “cheating” by conspiring with each other outside of recognized venues.

Another option might be to openly permit broad cross-media interaction, but focus data collection on key venues such as forums and specific physical locations, supplemented by embedded observers within the player community. This produces a data set that represents a small but meaningful sample of all game-related communications. However, it also risks major interactions gravitating to unrecorded venues.

A third option could be to create the game such that it can *only* be played with an emphasis on dedicated accounts or devices. This is the approach that has thus far been taken by *Ingress*. *Ingress* players interact through a variety of media, including private e-mails and in-person interactions.

However, the central activities of play require the game app, which among other things conveniently gathers geolocational data. Google has also encouraged communication through proprietary venues such as Google+.

Most human subjects studies follow a default of recording identity for purposes of consent and payment/credit, but storing it separately from all other data—but co-participants generally share information according to ordinary social norms. Communications through dedicated apps and programs open the possibility for ARG players to communicate and act anonymously or pseudonymously. Anonymity increases privacy, but also increases opportunities for participants to act unethically towards each other with minimal consequences (e.g., online bullying). Pseudonyms offer a compromise, allowing the development of in-game reputations separate from everyday identities. However, both these options become less possible as game play is integrated with regularly used online tools such as e-mail.

Game Boundaries. Many ARGs blur the boundaries between in-game and out-of-game settings, raising concerns about intersections between game-related activities and bystanders who have not consented to involvement. Traditional field studies may have similar issues, but are more likely to get consent from communities or organizations as a whole to facilitate their observation—a tactic that is likely to be possible for some research ARGs, but may not always be appropriate.

Bystander participation may range from observation to joining player activities, from lack of awareness that they have seen anything abnormal to full understanding of the game's nature. They may appreciate the encounter with something unusual, or be disturbed by it. Ethical guidelines have been suggested for some of these situations, but others can be more difficult (Montola, Stenros, & Waern, 2009). A few of Montola et al.'s examples may better illustrate this issue.

- Players in *Momentum* were told to get information from an NPC who would sit in a “mystical pattern” drawn in a public park. However, the players were delayed, the actor gave up and left, and an uninvolved homeless woman took her spot. Players assumed the homeless woman was their contact, and proceeded to search her possessions and attempt an in-game exorcism ritual, in the hopes that she would produce their looked-for clue.
- Also in *Momentum*, game-runners snuck their own painting into a local gallery; players then had to “case” the gallery and retrieve it. Gallery workers observed the players’ strange behavior, and had fun speculating about what was going on.
- Students in a game design class decided to put together an ARG. Their “rabbit holes” included scribbling on walls and oddly stacked chairs. When investigated, these were intended to lead to a story about a poltergeist. Instead, they were seen as an obnoxious act of vandalism by the janitorial staff who cleaned them up.

Other boundary-blurring situations might include outsiders attempting to join in clearly playful activity, family members being inconvenienced by time devoted to ARG-related activities, or “official” players inviting friends to help them solve puzzles without going through the full consent process.

Some of these non-players benefit from their experience, and some are inconvenienced or harmed—but ethical treatment does not guarantee benefit, and vice versa. For example, family may be inconvenienced by participation in a traditional lab study, but at least in the U.S., this is generally considered to be an acceptable choice on the part of the research participant. (Anthropologists working in more collectivist cultures sometimes do require familial consent, something that would be seen as unacceptable for a fully cognizant adult participant in most individualist cultures.) Players who have fun exploring puzzles, but who do not know

that they are contributing to research data, have not received an appropriate opportunity for informed consent.

Some of these concerns can be mitigated by more thoughtful preparation on the part of game-runners. If game play is to take place on private or institutional property, soliciting permission from appropriate authorities is a wise practice. Markers to distinguish game participants in public areas may also be helpful—ideally something that players can wear, rather than something that can be left behind for others to sit in (as in *Momentum*, above). Similar markers can be used for online sites, (e.g., the logo included on sites for the Smithsonian Institute’s *Ghosts of a Chance* game). These indicators are informative, non-intrusive, and have been found not to interfere with the ARG experience (Goodlander, 2009). However, even these best practices carry a certain degree of risk: property owners may neglect to inform employees of upcoming activities, and bystanders might coincidentally wear items that can be mistaken for ARG markers. At some point, eliminating all risk is likely to be impossible. Even in more traditional studies, researchers are not expected to prevent all possible adverse events, but rather take sensible precautions and respond appropriately when necessary.

Due to the complex, boundary-blurring nature of ARGs, some ethical responsibilities are likely to fall on the players as well as the organizers. Organizers may facilitate player ethics by providing appropriate guidance, tools, and processes, presumably as part of informed consent as well as available throughout the game. If players wish to recruit their friends to join in a game/study, they should be given easy ways to direct them to the consent process, and instructed to use this route. Clear cut ways of resolving ambiguity about possible participants could also be provided—for example, the aforementioned puppetmaster hotline, or passcodes that confirm game participation at the start of an interaction.

Less explicit markers can also be used to reassure bystanders. *The Beast's* many websites, associated with a wide range of fictional organizations, were not specifically labeled as game-related. However, they were prominently marked with the future dates on which they were ostensibly created or updated, ensuring that anyone who stumbled on them would recognize their fictional nature. Museum visitors acting strangely in jeans and black t-shirts might be viewed as potential criminals; museum visitors acting strangely in orange suits with purple polka dots, or t-shirts reading “Experimental Shakespeare Company,” are more likely to create the assumption that they are engaged in performance art.

Embedded, Independent Oversight. Generally speaking, the more novel, complex, and socially realistic a BSSR study is, the stronger the requirements for ethical oversight. As compared with more traditional lab-based methods, these types of studies carry greater risk of strong emotional responses, more opportunity for unexpected participant behavior, and more danger of missing potential adverse events. Just as the ethical review requirements are stronger for a medical study with physical risks, ARG-based research may also require more stringent oversight.

In addition to approving research prior to its implementation, Institutional Review Boards may also wish to require observers who can continue to evaluate ARG-based research activities while they are being carried out. This may include the ability to pause or stop activities that are perceived as problematic, or force withdrawal of participants who appear at risk.

Inadvertent Findings. The need for confidentiality and anonymity of participant information is an issue addressed by many BSSR studies. However, for any study producing a large, complex data set, additional ethical concerns may arise at the analysis level. For example, geolocation data can give unintentional information about participants’ work, home, and recreational habits—and can be readily used to re-identify

participants who were intended to be anonymous or pseudonymous. Other inadvertent findings—such as information that is not obviously available in the data set, but can be reconstructed from it—should also be anticipated. Researchers should consider ways to keep these data confidential, including strategies to prevent re-identification if analysis is crowdsourced.

When there is a risk of collecting information that could be compromising or stigmatizing, researchers are generally required to give assurances that this information will not be shared. This can range from simple confidentiality (e.g., when studying legal but denigrated subcultures), to requesting immunity from subpoenas (e.g., when studying medical error).

A particularly serious possibility is the potential for researchers in some studies to identify participants at risk of harming themselves or others. Under these circumstances, researchers are generally required to report, and to include that requirement in their informed consent documents. A related possibility is that of discovering risks that a participant was unaware of—for example, a study including a heart rate monitor might identify a potential heart murmur. Institutional Review Boards generally determine how such cases should be handled on the basis of experimenter expertise, and how strong these indications are likely to be. Experimenters who are not qualified to diagnose may not be permitted to report such indicators to participants.

Models for ARG-Based Research Ethics

When we think about alternate reality games and research ethics, the Stanford Prison Experiment (Zimbardo, n.d.) may not be an obvious example. However, there are similarities, and they provide potentially important lessons to keep in mind (Cutforth-Young, 2013).

In Zimbardo's classic and controversial experiment, young male college students signed up for a study (the rabbit hole), then were publicly “ar-

rested” and brought to a “jail” in the basement of the Stanford psychology building. There these new “prisoners” met other participants who had been given the role of guards. Participants fell into these roles, with prisoners seeking ways to break rules and guards abusing their power. This process was exacerbated by Zimbardo taking the role of prison superintendent, supporting the guards in their antisocial behavior and minimizing the degree to which that behavior was apparent to visitors.

The role-playing, the intrusion of the arrest procedure into participants’ everyday lives, and the immersive nature of the prison environment are similar to aspects of pervasive games. The complexity of the social interactions studied, and the data collected, are similar to those of interest in ARG-based research—although the methods and statistics of the time were not actually sufficiently advanced to appropriately collect or analyze that data. And the ethical issues highlighted by Zimbardo’s errors can be said to provide an almost perfect negative model:

- Informed consent did not include information that a reasonable person would want to know before participating. There were no opportunities for renegotiation during the experiment, and participants who attempted to exercise their right to quit were discouraged and belittled.
- All experimenters (Zimbardo and his associates) took active authority roles within the prison. While embedded observers are an important aspect of game-running and data collection, part of the team should remain focused on directing the overall experience. Furthermore, embedding observers only among player groups with greater power caused Zimbardo to miss a vital part of participant experience—and facilitated his failure to respond to that experience with the urgency it deserved. Experimenters should not take sides among potentially competing participant groups.

- Independent oversight was entirely absent. The experiment was eventually stopped through the intervention of one of Zimbardo's students; no standardized procedure was in place to meet this need.

The Stanford experiment provides a particularly interesting model because it was complemented, years later, by a BBC-sponsored partial replication that illustrates an ethical gold standard (Reicher & Haslam, 2006; Haslam & Reicher, n.d.):

- Informed consent was detailed, and opportunities for discussing issues available throughout the experiment.
- Data collection, and observation of participants for potential problems, was thorough (unsurprising, given the budget and resources of a television station).
- Experimenters were embedded in the prison environment, but also acted as separate observers. Independent ethicists and clinical psychologists also observed, retaining the right to pull participants for evaluation, remove participants from the experiment, or end the entire procedure at any time.

The BBC study was also a better controlled experiment—where Zimbardo essentially created a single social situation and let it run for field observation purposes, the BBC researchers included manipulation of independent variables and careful consideration of quantitative as well as qualitative measures. All of these factors combined to give them a more accurate and complete picture of social and psychological responses to an imposed hierarchy, and demonstrated that even while studying social complexity, good design and good ethics are compatible and complementary. Such comparisons should lend researchers some confidence that meaningful research with ARGs can be conducted without any compromise of ethical standards or best practices.

Risks and Mitigation

In order to create ARG-based research that is well-designed and ecologically valid, researchers are likely to wish to address a few key challenges. These overarching concerns have the potential to undermine entire experiments if not dealt with appropriately—and each represents an issue that should be considered at every stage of conception, design, and implementation.

Balancing Art and Science. For applied ARGs, prioritizing among multiple goals can seem like a daunting task. At times, artistic goals—a compelling story, an entertaining experience—may be interpreted as necessarily subordinate to more “practical” ends. However, there are a number of reasons why researchers will wish to resist this temptation.

The artistry that goes into an effective and entertaining ARG provides much of the power that improves ecological validity. Players may be more likely to respond in a psychologically realistic manner, more reflective of real-world behavior, should they find themselves in a highly transporting narrative. A compelling storyline and interesting challenges are also what draw players into an ARG, and maintain their interest and participation over extended periods of time. As an optional activity available to the public, a research ARG must compete with many intriguing and entertaining possibilities. Even ARGs that focus on a particular organization, where players may initially join because participation is required or encouraged, are likely to benefit from immersing players in a compelling situation.

However, there may still be times when art and science conflict. Ordinary ARGs may plan only the general outlines of a story, and remain extremely open to being shaped by player actions and interests. However, both for design reasons and because of human subjects review requirements, research ARGs may require a more detailed or stable plan. Find-

ing the balance between these needs may take some trial and error. Poor design could also lead to situations in which a compelling story does not support the desired research questions.

One option for mitigating these concerns is to carefully plan research and art as an organic whole—for example, by explicitly designing stories around specific research questions, and ensuring narrative justification for experimenter manipulations. Game-runners should also gather feedback on NPCs to ensure that they are perceived as psychologically realistic and believable entities. This step may be particularly important when working with, and/or creating characters who represent, a cross-cultural sample.

Replication and Scaling. Replication is a cornerstone of scientific research, and is becoming particularly important in the social sciences, where rigorous replication efforts have previously been lacking (Klein et al., 2014). However, ARG-based research is likely to present replication challenges if judged by the standards of lab research. While precise game design and use of manipulations may be repeatable, no participant's actual experience will be. Those experiences are influenced by the intersection of research design and the behavior of other participants. For this reason, ARG-based research replication may be more like replication of field observations, in which researchers seek out (or in this case create) similar circumstances, and look to see whether outcomes are also similar—although ARG researchers will have the advantage of being able to recreate specific narratives and independent variables.

Hansen et al. (2012) provides a careful examination of this topic, albeit from the perspective of game designers wishing to increase their return on investment rather than researchers wishing to examine the same hypothesis repeatedly. Spoilers (in research terms, difficulty finding naïve subjects) are a particular threat to replicable ARGs: popular games are extensively documented and discussed on online message boards, making it difficult to prevent later participants from learning about key plot

points, goals, and characters. The authors suggest a number of possible strategies to overcome this barrier, a few of which are particularly relevant to research ARGs:

- *Cyclical, Event-driven play.* This design style matches the game with a recurring event—ideally one that brings in a new, naïve cohort each time. This works well with local, focused populations. For example, ARGOSI and Reality Ends Here, both intended as university orientation activities, recreate a similar experience for each incoming class of freshmen.
- *Secret Player Community.* Players have access to a private portal where material and discussions are made available to current players, but not the general populace. Groups can also be kept separate, and information-sharing discouraged, through competition—although this may well push communications to private media that are hard for researchers to access.
- *Ongoing Exhibit.* Play takes place at a specific location, and individual or groups can experience them only when they are present. The modular version of the Smithsonian’s Ghosts of a chance, or Colonial Williamsburg’s Revquest, are examples. Potential players are motivated to avoid too much information about the game, but information is likely to remain available for those who seek it out.
- *Multiple Seasons or episodes.* New rounds of a game build off of, but do not repeat, previously created material. This is not strictly a replication method, but is mentioned here because it provides a good model for testing iterative hypotheses. Games such as Heroes 360, that build their narratives around ongoing television shows, are good examples of this strategy. Revquest, which changes episodes periodically to attract new visitors, is another.

One variation, suggested by the “secret player community” strategy, is the possibility of running identical or similar research ARGs in parallel. Either a single team, or two teams working in concert, could test the same

hypotheses, using similar games, across different locations or populations. While keeping play synchronized between tracks might be a challenge, it could potentially provide a fast metric for robustness through comparison of player responses.

Another challenge to replication, particularly across labs, is recording and preserving original methods and materials (Hansen et al., 2012). The transmedia nature of ARGs, and the complexity of interactions that go into them, makes them particularly challenging to archive at the appropriate level of detail. Game organizers who care about this issue should plan for archival records in advance. The Electronic Literature Organization and the Stanford Humanities lab have worked to address these issues, but standard archival techniques tend to tie together materials with a single domain name—a problematic practice for ARGs.

Behavioral Validity. While one of the most promising aspects of ARG-based research is the potential for highly ecologically valid situations, players themselves may interfere with this goal. For the players who immerse themselves most deeply in a fictional scenario, one appeal may be the opportunity to play a role or even a personality separate from their everyday self. Games provide a “safe space” to experiment with alternate selves, and try out more heroic or antisocial behavior than one’s usual tendencies.

Some of these tendencies may be encouraged by the use of particular narratives—for example, players inducted into a secret group of spies may respond very differently than those inducted into a secret group of superheroes, even if both are faced with the sudden onset of a zombie uprising. While Zimbardo’s findings from the Stanford prison experiment have since been examined with more nuance (Reicher & Haslam, 2006), roles and situations do have a significant effect on behavior. Even a neutral ‘mainstream’ narrative will be subject to this—it is very hard to create a story that does not invoke some scripts for how to respond to

it. Researchers may wish to think carefully about these scripts, and take them into account when drawing conclusions from their results.

Participants in a research ARG can be encouraged to “play themselves” by using real names (which raises its own issues as discussed above, but may be acceptable if treated carefully). Montola, Stenros, & Waern (2009) suggest that players act more like themselves when objects and actions represent themselves in the game. That is, players are more likely to take on new roles when a stick stands in for a sword than when a realistic-looking toy sword stands in for a sword, or when players are asked to actually pick a lock rather than solving a puzzle on their smart phones to achieve the same affect. Sensory realism can cue psychological realism.

Emergent Player Behaviors. ARG players frequently push puppetmaster-set boundaries on possible activities and the shape of the narrative. Experienced game designers often consider themselves to be in collaboration with their participatory audience, and play with the tension between the multiple creative loci. While research goals may limit the degree to which this can be encouraged, a certain degree of player initiative must be taken into account—and for many players, that opportunity is part of the appeal.

How Can Research ARGs Implement Randomized Controlled Trials?

Randomized controlled trials (RCTs) are a gold standard for inferring causality. Interventions or other variable manipulations are randomly assigned to some individuals, while others in comparable situations are assigned to a control condition; differences are then assessed (Haynes, Service, Goldacre, & Torgerson, 2012).

However, when working in complex real world and near-real-world environments, designing and implementing an RCT can be challenging. Appropriate design of the control condition, the scale at which random assignment is to take place, the appropriate length of study, and ethical scaling of apparently successful effective interventions, are all potentially contentious topics.

The increased ability to control extraneous variables in ARG-based research may provide an opportunity for improved RCTs. Unlike a field study, participants can be assigned at random to teams or factions. However, defining control conditions, and responding appropriately when one condition appears to confer outsized benefit or risk, will remain important and difficult considerations.

Emergent behaviors may range from the deliberate disruptions experienced by *Conspiracy for Good* to the prosocial creation of tools for new recruits common among habitual ARG players. In between these two extremes, players speculate about—and respond to—puppetmaster plans that bear no resemblance to any actual intentions. They mistake unrelated materials for game-based ones, or irrelevant aspects of game materials for vital clues (Phillips, 2012). They cooperate with in-game rivals in order to make the plot more interesting, or just for fun. Ingress players regularly neglect the deadly battle between Enlightened and Resistance

in order to create country-spanning art with their blue and green portal claims (e.g., Dhaider, 2013). Some of these unexpected behaviors may be disruptive to data collection; others may turn out to be the equivalent of Alexander Fleming's mold-infested petri dishes.

Making a Difference: How ARG-Based Methodologies Can Improve BSSR Research

With this background, we now focus on some of the specific advantages offered by ARG-based research, and research topics that might be most appropriate for initial exploration. We have referred to these ideas throughout the book, but bring them together here.

How ARG-Based Research Can Improve on Current Methods

Carefully designed and implemented ARGs can be used to explore behavioral and social science topics in ways not available through other methodologies. Both lab and field studies have their strengths as well, and have long been seen as complementary. Mixed methods, which combine some experimental components with richer field settings, are an underutilized but vital addition to that portfolio (Grant & Wall, 2009). ARG-based research may be seen as a special case that enhances the potential of mixed methods and helps develop a richer understanding of social and behavioral interactions.

Perhaps the most important potential of ARG-based research lies in the ability to study complex behaviors in complex social contexts, but with rich opportunities for controlled experimentation over shorter or longer periods of time. By combining observational techniques with manipulation of independent variables, researchers can produce detailed data sets that are amenable to new analytic methods, and causal as well as descriptive conclusions about behavioral interaction. More traditional mixed methods are able to examine specific independent variables, but have a

limited ability to control extraneous variables. While the ability to do this in ARGs does not match the level of control available in the lab, it is likely to be superior to what has previously been available in naturalistic settings.

ARGs—as entertaining, interactive activities based in rich narratives that are overlaid upon everyday experience—may be far more engaging for participants than most research methods. ARGs draw a relatively diverse population, and even many “small” ARGs easily recruit large numbers of participants by behavioral study standards. ARGs can take advantage of outreach methods that are not normally good for recruiting research participants, such as viral marketing and word of mouth. Narratives can be tailored to attract and retain populations of interest, or players best suited to address particular questions. Alternatively, an ARG or Alternate Reality Environment with an open-ended narrative could be used to engage long-term participants who are available at need to explore and address urgent questions.

Where Should We Start?

ARG-based research opens many opportunities for examining BSSR topics in ways not previously possible. However, it is also a largely untested area: reliability and validity remain to be demonstrated, and the best designs have not yet been identified. Choices about pilot topics and designs will heavily influence the scientific community’s acceptance, or lack thereof, of these new methodologies. The field will benefit from a mix of efforts that replicate and extend well-studied phenomena with those that demonstrate ARGs’ potential for examining previously difficult areas.

Well-replicated findings, particularly those that have remained consistent across lab and field studies, can be used to test the essential validity of ARG-based research. Findings from industrial/organizational psychology have been particularly robust (Mitchell, 2012) and are well-suited to many ARG designs. Topics such as team dynamics, motivation to com-

plete tasks, and leadership style have already been studied observationally in ARGs that were not explicitly designed to permit that study (e.g., Gurzick et al., 2011; Peyton, Young, & Lutters, 2013). Recent efforts to plan parallel replication attempts across labs (e.g., Klein et al., 2014) could be extended to include ARG studies. This would provide proof of principle for areas in which ARG-based research can duplicate and extend robust results, and potentially also provide indicators of appropriate follow-ups.

There are many research topics that have not been amenable to previously existing research methods, or that could be explored in new ways. These may include:

- *Phenomena that would benefit from testing in more complex or controlled contexts.* Cognitive and social neuroscience labs produce a wide range of findings under heavily controlled circumstances, some of which may be masked or otherwise difficult to observe in field studies. ARGs could provide a venue for examining whether, and how, these phenomena play out in the context of realistic social and physical interaction. For example, sacred values have been identified in the lab that may have overweighted effects on behavior. These values are both heavily context-driven, and difficult to study in a controlled fashion in the field—a research ARG might improve understanding of how they work in situ.
- *Social science phenomena that emerge only in large groups, in complex interactions between groups, or over extended periods of time.* Hypotheses emerging from field research may benefit from exploration under more controlled circumstances. For example, ARGs that divide players into factions could be used to examine the emergence of clandestine communications among competing groups or anti-social behavior within a group.
- *Phenomena that emerge only during unexpected or emergency situations.* Immersive environments and war games have long been used to give

emergency responders opportunities for realistic practice. This is also the focus of existing ARG-based research (Fischer et al., 2012). Future research could explore spontaneous collaboration in crisis situations, or the influence of crowdsourced information distribution.

- *Effects of multiple communication media.* Communications through different media, and with different degrees of presence, are central to the ARG experience—making it the perfect venue for testing these variables (Gurzick et al., 2011). The creation of effective virtual teams poses a challenge to many modern organizations, and ARGs often feature such teams interacting successfully to produce high-quality work.
- *Phenomena based in research with small samples.* Fields such as neuroscience often derive hypotheses or conclusions from research with small Ns due to technical or financial limitations. While some of these areas remain dependent on time-consuming processes and expensive equipment, new portable sensors (e.g., wearable EEG devices) permit at least preliminary studies in more realistic, contextual settings.
- *Small but important effects.* Many rare phenomena, or phenomena with small effect sizes, are extremely meaningful from a practical standpoint. Examples might include positive deviance—unusually altruistic behavior—as well as rare but dangerous antisocial phenomena (e.g., the emergence of radicalism or anti-social behaviors in new communities). Methodologies that bring in large numbers of participants over an extended time period may make these easier to detect and study in depth.

Money, Time, Metrics

Good ARG-based research is likely to require planning based on realistic expectations for costs and logistics, as well as clear measures. Based on current evidence, a wide range of options may be possible for each of these.

Budgetary considerations. As described above, budgets for ARGs vary tremendously. Although the product of a \$100 investment looks little like what can be produced with a \$4 million budget, the relationship between investment and quality is not linear. An experienced development team can create an effective experience at any reasonable price point. Budget level does, however, have a strong effect on how much content can be created and offered, and how quickly (IGDA, 2006). This in turn influences player recruitment and retention. A research game that requires a large N may wish to invest additional funds in content creation, outreach efforts, and staff to ensure responsiveness during player interactions. Budget also influences the ability to organize complex live events or produce tangible artifacts.

Are Players Willing to Pay?

Given that ARGs are an entertaining experience that requires skilled experts to produce, it seems sensible to many designers that people might pay to participate. In practice, this type of funding has not been easy to implement.

Majestic was the first attempt to create a pay-to-play ARG. Produced by a traditional video game company, it was canceled after less than a year due to low sign-ups. *Perplex City* and *Edoc Laundry* were more successfully funded by product sales, but neither was successful enough to remain viable long-term.

Some short ARGs taking place in controlled environments have more successfully charged for admission. *Goruck Challenges* charge for a day-long physical or intellectual intensive experience, and location-based games such as *Spy in the City* have successfully produced revenue over extended periods of time.

It is not certain why longer player-funded ARGs have not been successful. The precedent and ongoing availability of free-to-play games may reduce interest, and the distributed nature of ARG play may make it difficult to explain precisely what people are paying for. For now, other funding models remain more consistently viable.

For a game with a research component, budget will also be needed for additional staff with topical expertise, resources and time for analysis, and specialized equipment. A larger budget will permit researchers to acquire portable sensors or dedicated devices for player use, support participation in multiple locations, or explore rich data sets in greater depth.

Researchers interested in planning ARG-based research may want to consider the following line items:

- Staff time for the various experts listed under “Best Practices in Staffing.” Many experienced ARG designers work as freelancers or as part of creative services companies; fees and pay structures vary considerably. Staff will be needed during planning, implementation, and follow-up/analysis stages.
- Website hosting and creation of game-related sites
- Production budget for video components
- Dedicated devices or online services for player communication
- Sensors and other data collection tools
- Server space and appropriate programs for analyzing large data sets.
- Outreach efforts and materials
- Creation of in-game artifacts and physical puzzles
- Materials for live events
- Participant payments. Although playing a well-designed game should be intrinsically motivating, some IRBs may require payment for research activities. In addition, even enjoyable ARGs can be time-consuming and inconvenient, and payment may ease logistics for players to participate more fully and often. Games that do not pay participants may still wish to offer financial aid with travel to live events, dedicated communication devices that can be kept following play, or other rewards and incentives.

Time Frame. Successful ARGs have been carried out over the course

of hours, years, and everything in between. Specific time frames may be chosen on the basis of the research topic of interest—or based on whether a game is intended to study a single question or encompass multiple iterative studies. Narratives and social interactions may play out differently over time.

Many extremely short ARGs tend to be designed for demonstration purposes, as intensive publicity bursts, or as semi-casual recreational opportunities. Many of these are local (e.g., *Survive DC*, or events during *Come Out and Play* festivals) and expand activities such as scavenger hunts to include a narrative component and a wider range of problem solving activities. Foresight Engine games are also brief, typically running for under 72 hours. Short games may be suitable for attempting to replicate lab-based phenomena, for quick pilot tests of new hypotheses, or for crowdsourced hypothesis generation around a new question.

The longest ARGs tend to be heavily sponsored, often marketing or expanding larger franchises. *Perplex City* ran for 2 years, funded by sales of puzzle cards, while television shows such as *Heroes* have run games in tandem with multiple seasons. While expensive and logistically challenging, these games may eventually prove extremely effective for simulating and exploring the development of social dynamics over time—the emergence of factions, large-scale attitude shifts, or factors that influence the long-term success of negotiated agreements, for example.

Many ARG-like activities, however, do not consist of one time-limited story. Live action role-playing communities, for example, may persist for many years, creating new storylines and interactions. Transmedia franchises also maintain their worlds and associated communities over long periods of time. Given sufficient proof of concept from shorter duration projects, it may be possible to create research ARGs that act as ongoing resources—much as universities currently maintain labs and subject pools rather than reconstructing them for each individual experiment. These

communities have extensive turnover, but maintain coherent cultures and practices. An ongoing research ARG community might have a range of ongoing storylines, teams, and factions, designed in such a way that new research questions could be created as modules and inserted into play. Such a set up would carry risks—for example, insufficient turnover could lead to a homogenous subject pool that was overly well-informed about current psychological theory. Nevertheless, the idea could be worth exploring in further detail.

Metrics. Two types of metrics are of concern for ARG-based research design. First, measures will be necessary that directly address research questions. Second, other measures will focus on the success of the game, examining outreach and participation levels.

With regard to research measures, ARGs offer a wide range of possibilities, discussed in the section on data collection. Designers might wish to consider which will be most informative for their research question, easiest to fit into the game's narrative, and most suitable for the available media. They may also consider their analytic capabilities.

Metrics of overall success are likely to focus on factors that directly or indirectly influence research goals, including:

- *Number of active players.* It may also be relevant to measure number of initial sign-ups or registrations, or participants who start playing the game but don't complete it—particularly to determine what percentage of total involvement is made up by active players, and to direct future outreach and retention efforts. However, many ARGs focus on these larger figures when reporting metrics, masking the number of people who engage in meaningful interaction throughout the course of the project. The active players who complete play are those whose data will actually be available for analysis, and should not be neglected.

- *Participation levels.* Even players who are active throughout the course of a game may not participate in all activities. Researchers may wish to know which activities were most or least engaging, and how much time different participants devoted to various activities.
- *Outreach success.* People who hear about the game but don't play, or people who remain passive observers, may nevertheless contribute to public attitudes toward the research, or pass information on to those who do become active players. Measures of social media sharing, or positive and negative press coverage, can help build a better picture of these broader attitudes and interpretations.

Chapter 10

Conclusions

Alternate reality games have great potential to increase the range of methods and research questions available to the behavioral and social science community. They provide a framework that can be used to study social phenomena in a complex, near-real-world environment, with high levels of experimenter control, and could substantively advance efforts at new forms of quantitative human science. However, many questions remain to be addressed, and much work will likely be needed to overcome the challenges of developing a novel methodology. There is some risk that, eventually, ARG-based research may prove to have barriers that overwhelm its possible advantages. Nevertheless, those advantages are impressive enough, and the likelihood of success high enough, to make the attempt worthwhile.

Behavioral and social science research stands poised to enter a new era—one of more rigorous experimentation and replication, and deeper and more extensive analysis, than has previously been possible. Along the way, new methods will be necessary to fill in gaps and resolve ambiguities from conflicting findings. Alternate reality games offer a potentially powerful addition to the portfolio of research tools used by behavioral and social scientists.

Bibliography

- Andersen, M. (2011). A behind-the-scenes look at the making of Cisco's ARG *The Hunt*. *Wired*. Retrieved from: <http://www.wired.com/2011/01/a-behind-the-scenes-look-at-the-making-of-ciscos-game-the-hunt/>
- Audubon. (2014). Christmas bird count. Retrieved from: <http://birds.audubon.org/christmas-bird-count>
- Bainbridge, W. S. (2007). The scientific research potential of virtual worlds. *Science*, *317*, 472-476.
- Balakrishnan, B., & Sundar, S. S. (2011). Where am I? How can I get there? Impact of navigability and narrative transportation on spatial presence. *Human-Computer Interaction*, *26*, 161-204.
- Ball, P. (2014). Strength in numbers. *Nature*, *506*, 422-423.
- Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep: Some answers to criticisms of laboratory experiments. *American Psychologist*, *37*, 245-257.
- Besmann, A., & Rios, K. (2012). Pals in power armor: Attribution of human-like emotions to video game characters in an ingroup/outgroup situation. *Cyberpsychology, Behavior, and Social Networking*, *15*, 441-443.
- Bonsignore, E., Hansen, D., Kraus, K., & Ruppel, M. (2012). Alternate reality games as platforms for practicing 21st-century literacies. *International Journal of Learning*, *4*, 25-54.
- Bonsignore, E., Kraus, K., Visconti, A., Hansen, D., Fraistat, A., & Druin, A. (2012, May). Game design for promoting counterfactual thinking. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* (pp. 2079-2082). ACM.
- Brandt, D. J., Sommer, J., Krach, S., Bedenbender, J., Kircher, T. Paulus, F.M., & Jansen, A. (2013). Test-retest reliability of fMRI brain activity during memory encoding. *Frontiers in Psychiatry*, *4*, 1-9

- Borden, J. D. (2012). A.R.G. – Creating alternative reality games for the classroom. Retrieved from: http://sloanconsortium.org/effective_practices/arg-creating-alternative-reality-games-classroom
- Boskik, N. (2011, October). Ethical behavior, personal growth and becoming a citizen of the world: The power of online games. In *Association of Internet Researchers 12.0: Performance and Participation*.
- Brouillette, L. (2009). How the arts help children to create healthy social scripts: Exploring the perceptions of elementary teachers. *Arts Education Policy Review*, 111, 16-24.
- Brown, C. R., Watkins, A., & Greitzer, F. L. (2013, January). Predicting insider threat risks through linguistic analysis of electronic communication. In *2013 46th Hawaii International Conference on System Sciences (HICSS)* (pp. 1849-1858). IEEE.
- Bunting, B. S., Hughes, J., & Hetland, T. (2012). The player as author: Exploring the effects of mobile gaming and the location-aware interface on storytelling. *Future Internet*, 4, 142-160.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14, 365-376.
- Center for Disease Control. (2012, September). Zombie preparedness. Retrieved from: <http://www.cdc.gov/phpr/zombies.htm>
- Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science*, 9, 40-48.
- Christian, B. (April 25, 2012). The A/B test: Inside the technology that's changing the rules of business. *Wired.com*. Retrieved from: http://www.wired.com/2012/04/ff_abtesting/
- Connolly, T. M., Stansfield, M., & Hainey, T. (2011). An alternate reality game for language learning: ARGuing for multilingual motivation. *Computers and Education*, 57, 1389-1415.
- Cutforth-Young, C. (2013). Spies like me: My response to IARPA's RFI UAREHERE. Retrieved from: <http://queenspade.com/my-open-re>

sponse-to-iarpas-rfi-uarehere/

- Davis, J. I., & Markman, A. B. (2012). Embodied cognition as a practical paradigm: Introduction to the topic, the future of embodied cognition. *Topics in Cognitive Science, 4*, 685-691.
- Defense Advanced Research Projects Agency. (2009). DARPA Network Challenge. Retrieved from <http://archive.darpa.mil/networkchallenge/>
- Defense Advanced Research Projects Agency. (n.d.). Narrative Networks. Retrieved from: http://www.darpa.mil/Our_Work/DSO/Programs/Narrative_Networks.aspx
- Dena, C. (2008a). ARG stats. Retrieved from <http://www.christydena.com/online-essays/arg-stats/>
- Dena, C. (2008b). Emerging participatory culture practices: Player-created tiers in alternate reality games. *Convergence: The International Journal of Research Into New Media Technologies, 14*, 41-57.
- Dhaider. (2013). Cross faction Ingress art. Retrieved from: <http://www.youtube.com/watch?v=NiLoODbo43w>
- Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*, 274-290.
- Disney Imagineering Research and Development. (n.d.). Living worlds: Bringing immersive stories to life in the real world. Retrieved from: <http://disneylivingworlds.com/>
- DTRA. (2013). DTRA/SCC-WMD announces \$1 million algorithm challenge winner. Retrieved from: <http://www.prweb.com/releases/dtra/algorithmchallenge/prweb11150852.htm>
- Efron, B. (2005). Modern science and the Bayesian-frequentist controversy. In *Papers of the XII Congress of the Portuguese Statistical Society*, 9-20.
- Ekman, I., Chanel, G., Jarvela, S., Kivikangas, J. M., Salminen, M., & Ravaja, N. (2012). Social interaction in games: Measuring physiological linkage and social presence. *Simulation Gaming, 43*, 321-338.
- Ermi, L., & Mäyrä, F. (2005, May). Fundamental components of the

- gameplay experience: Analyzing immersion. In *Proceedings of DiGRA 2005 Conference: Changing Views—Worlds in Play*.
- Fischer, J. E., Flintham, M., Price, D., Goulding, J., Pantidi, N., & Rodden, T. (2012). Serious mixed reality games. In *Mixed Reality games: Workshop at the 2012 ACM Conference on Computer Supported Cooperative Work*.
- Fischer, J. E., Jiang, W., Kerne, A., Greenhalgh, C., Ramchurn, S. D., Reece, S., Pantidi, N., & Rodden, T. (2014). Supporting team coordination on the ground: Requirements from a mixed reality game. In *Proc. 11th Int. Conference on the Design of Cooperative Systems (COOP '14)*. Springer.
- Fischer, J. E., Jiang, W., & Moran, S. (2012). AtomicOrchid: a mixed reality game to investigate coordination in disaster response. In *Entertainment Computing-ICEC 2012* (pp. 572-577). Springer Berlin Heidelberg.
- Fox, J. (2011, October 11). Frequentist vs. Bayesian statistics: resources to help you choose (updated). Retrieved from: <http://oikosjournal.wordpress.com/2011/10/11/frequentist-vs-bayesian-statistics-resources-to-help-you-choose/>
- Fulmer, C. A., & Gelfand, M. J. (2012). At what level (and in whom) we trust: Trust across multiple organizational levels. *Journal of Management*, 38, 1167-1230.
- Garrand, T. (1997). Scripting narrative for interactive multimedia. *Journal of Film and Video*, 49, 66-79.
- Gerrig, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. New Haven, CT: Yale University Press.
- Goodlander, G. (2009). Fictional press releases and fake artifacts: How the Smithsonian American Art Museum is letting game players redefine the rules. In J. Trant and D. Bearman (eds). *Museums and the Web 2009: Proceedings*. Toronto: Archives & Museum Informatics. Retrieved from: <http://www.archimuse.com/mw2009/papers/goodlander/goodlander.html>
- Goruck. (2014). Goruck challenge. Retrieved from: <https://www.goruck.com>

com/challenge

- Grant, A. M., & Wall, T. D. (2009). The neglected science and art of quasi-experimentation: Why-to, when-to, and how-to advice for organizational researchers. *Organizational Research methods*, 12, 653-686.
- Green, M. C., & Brock, T. C. (2002). In the mind's eye: Transportation-imagery model of narrative persuasion. In Green, M. C., & Strange, J. J. (Eds.) *Narrative impact: Social and cognitive foundations* (pp. 315-341). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Greiner, B., Caravella, M., & Roth, A. E. (2014). Is avatar-to-avatar communication as effective as face-to-face communication?—An ultimatum game experiment in first and Second Life. *Journal of Economic Behavior & Organization*.
- Gurzick, D., White, K. F., Lutters, W. G., Landry, B. M., Dombrowski, C., & Kim, J. Y. (2011, February). Designing the future of collaborative workplace systems: lessons learned from a comparison with alternate reality games. In *Proceedings of the 2011 iConference* (pp. 174-180). ACM.
- Hagan, J. (2012, July 26). EnergyMMOWGLI, Navy tackle energy challenges. Retrieved from: <http://www.iftf.org/future-now/article-detail/energymmowgli-helps-navy-tackle-energy-challenges/>
- Hansen, D., Bonsignore, E., Ruppel, M., Visconti, A., & Kraus, K. (2012). Designing reusable alternate reality games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1529-1538). ACM.
- Haslam, S. A., & Reicher, S. (n.d.). The BBC prison study. Retrieved from: <http://www.bbcprisonstudy.org/index.php>
- Haynes, L., Service, O., Goldacre, B., & Torgerson, D. (2012). Test, learn, adapt: Developing public policy with randomized controlled trials. *Cabinet Office: Behavioral Insights Team*.
- Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61-135.

- Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 58, P3-P11.
- Howard, G. S., Lau, M. Y., Maxwell, S. E., Venter, A., Lundy, R., & Sweeny, R. M. (2009). Do research literatures give correct answers? *Review of General Psychology*, 13, 116-121.
- International Game Developers Association. (2006). Alternate reality games white paper.
- International Game Developers Association. (n.d.). Alternate reality games SIG/whitepaper/types of ARG. Retrieved from: http://wiki.igda.org/Alternate_Reality_Games_SIG/Whitepaper/Types_of_ARG
- Institute for the Future. (n.d.). Engaged foresight in action. Retrieved from: <http://www.iff.org/what-we-do/foresight-tools/collaborative-forecasting-games/foresight-engine/>
- Intelligence Advanced Research Projects Activity. (2013, March). Request for information (RFI) – Using alternate reality environments to help enrich research efforts (UAREHERE). Retrieved from: http://www.iarpa.gov/RFI/rfi_uarehere.html
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2, e124.
- Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, 7, 645-654.
- Jensvald, J., & Morin, M. (2004). Simulation-supported live training for emergency response in hazardous environments. *Simulation and Gaming*, 35, 363-377.
- Kahneman, D. (2012, September 26). A proposal to deal with questions about priming effects. Retrieved from: http://www.nature.com/polopoly_fs/7.6716.1349271308!suppinfoFile/Kahneman%20Letter.pdf
- Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., & Eastwood, J. D. (2003). Attention, researchers! It is time to take a look at the

- real world. *Current Directions in Psychological Science*, 12, 176-180.
- Klein, R. A., Ratliff, K., Nosek, B. A., Vianello, M., Pilati, R., Devos, T., Galliani, E. M., et al., (2014). Investigating variation in replicability: The “many labs” replication project. Retrieved from Open Science Framework, <http://osf.io/wx7ck>
- Lombard, M., and Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer Mediated Communication*, 3, 0-0.
- Mapsurfer, T. (2005). Letterboxing in America. Retrieved from <http://www.letterboxing.org/faq/faq.html>
- McGonigal, J. (March 2003). A real little game: The performance of belief in pervasive play. Level Up Conference, Center for New Media: University of California.
- McGonigal, J. (2007). Why I Love Bees: A case study in collective intelligence gaming. *The ecology of games: Connecting youth, games, and learning*, 199-227.
- McGonigal, J. (2008). Chaotic community (updated). Retrieved from: <http://blog.avantgame.com/2008/06/chaotic-fiction-meet-chaotic-community.html>
- McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. New York NY: The Penguin Press.
- Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, 67, 371-378.
- Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science*, 7, 109-117.
- Montola, M., Stenros, J., & Waern, A. (2009). *Pervasive games: Theory and design*. Burlington, MA: Morgan Kaufmann Publishers.
- Moran, S., Pantidi, N., Bachour, K., Fischer, J. E., Flintham, M., Rodden, R., Evans, S., & Johnson, S. (2013, March). Team reactions to

- voiced agent instructions in a pervasive game. In *Proceedings of the 2013 international conference on intelligent user interfaces* (pp. 371-382). ACM.
- Morgan, S. E., Movius, L., & Cody, M. J. (2009). The power of narratives: The effect of entertainment television organ donation storylines on the attitudes, knowledge, and behaviors of donors and nondonors. *Journal of Communication*, *59*, 135-151.
- Moseley, A. (2008). An alternative reality for higher education? Lessons to be learned from online reality games. In ALT-C.
- Moseley, A., Alexander, B., & Bono, J. J. (2010). ARGs in education and training. *Argology*. Retrieved from: <http://www.argology.org/args-in-education-training/>
- Moseley, A., Whitton, N., Culver, J., & Piatt, K. (2009). Motivation in alternate reality gaming environments and implications for education. In *Proceedings of the 3rd European conference on game-based learning*.
- Mott, B. W., Callaway, C. B., Zettlemoyer, L. S., Lee, S. Y., & Lester, J. C. (1999, November). Towards narrative-centered learning environments. In *Proceedings of the 1999 AAAI Fall Symposium on Narrative Intelligence* (pp. 78-82).
- National Research Council. *New Directions in Assessing Performance Potential of Individuals and Groups: Workshop Summary*. Washington, DC: The National Academies Press, 2013.
- Nordic Larp Talks (n.d.) Retrieved from: <http://nordiclarptalks.org/>
- Nosek, B. A., & Lakens, D. (eds.) (2014). *Social Psychology*, *45*, Special Issue on replication.
- O'Hara, K., Grian, H., & Williams, J. (2008, December). Participation, collaboration and spectatorship in an alternate reality game. In *Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat* (pp. 130-139). ACM.
- OldWeather. (n.d.) Retrieved from: <http://www.oldweather.org/>

- Orne, M.T., (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776-783.
- Pashler, H., & Wagenmakers, E. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528-530.
- Peyton, T., Young, A. L., & Lutters, W. (2013, April). Playing with leadership and expertise: military tropes and teamwork in an arg. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 715-724). ACM.
- Phillips, A. (2012). *A creator's guide to transmedia storytelling*. New York NY: McGraw Hill.
- Pinchbeck, D. M., & Stevens, B. (2005). Presence, narrative and schemata. In *Proceedings of Presence 2005: The 8th International Workshop on Presence* (pp. 221-26).
- Reicher, S., & Haslam, S. A. (2006). Rethinking the psychology of tyranny: The BBC prison study. *British Journal of Social Psychology*, 45, 1-40.
- Schneier, B. (2012). *Liars and outliers: Enabling the trust that society needs to thrive*. Indianapolis IN: John Wiley and Sons.
- Smith, E. R., & Semin, G. R. (2007). Situated social cognition. *Current Directions in Psychological Science*, 16, 132-135.
- Smith, R. (2010). The long history of gaming in military training. *Simulation & Gaming*, 41, 6-19.
- Stark, L. (2012). *Leaving Mundania: Inside the transformative world of live action role-playing games*. Chicago, IL: Chicago Review Press.
- Stenros, J. (2008). The frame of pervasive live action role-playing games: A case study applying frame analysis on *Momentum*. Unpublished Masters thesis. Retrieved from: <http://stenros.kapsi.fi/pics/Stenros-2008FramesOfPervasiveLarp.pdf>
- Stenros, J., Holopainen, J., Waern, A., Montola, M., & Ollila, E. (2011). Narrative friction in alternate reality games: Design insights from

- Conspiracy For Good. In *Proceedings of DiGRA 2011 Conference: Think Design Play*.
- Stenros, J., Waern, A., & Montola, M. (2012). Studying the elusive experience in pervasive games. *Simulation Gaming*, 43, 339-355.
- Stokes, B., Watson, J., Fullerton, T., & Wiscombe, S. (2013). A reality game to cross disciplines: Fostering networks and collaboration. In *Proceedings of Digital Games Research Association 2013*.
- Szulborski, D. (2005). *This is not a game: A guide to alternate reality gaming*. New Fiction Publishing.
- Tester, J. (2013, September 27). Value of Research Vanguard 2038 award! Retrieved from: <http://blog.innovate2038game.org/value-of-research-vanguard-2038-award/>
- Twain, M. (1876). *The adventures of Tom Sawyer*. American Publishing Company.
- Unfiction. (n.d.). Unfiction alternate reality gaming. Retrieved from <http://www.unfiction.com/>
- USAID. (2013). Meet the tech challenge model winners. Retrieved from <http://thetechchallenge.org/winners/model.html>
- Volkova, S., Wilson, T., & Yarowsky, D. (2013, October). Exploring demographic language variations to improve multilingual sentiment analysis in social media. In *Proceedings of the 2013 Conference on Empirical Methods on Natural Language Processing*.
- Waern, A., Ahmet, Z., & Sundström, D. (2009, October). An in-game reporting tool for pervasive games. In *Proceedings of the International Conference on Advances in Computer Entertainment Technology* (pp. 240-248). ACM.
- Whitton, N. (2008). Alternate reality games for developing student autonomy and peer learning. *Proceedings of Learners in the Co-creation of Knowledge*, 32-40.
- Whitton, N. (2009). *Alternate reality games for orientation, socialization and induction (ARGOSI)*. Technical report, JISC.
- Yee, N., Bailenson, J. N., Urbanek, M., Chang, F., & Merget, D. (2007).

The unbearable likeness of being digital: The persistence of non-verbal social norms in online virtual environments. *Cyberpsychology and Behavior*, 10,115-121.

Young, A. L., & Lutters, W. (2013). The (not so) invisible hand: Perceptions of game designers and the impact on play. *Selected Papers of Internet Research*, 3.

Zimbardo, P. G. (n.d.). Stanford prison experiment: A simulation study of the psychology of imprisonment conducted at Stanford University. Retrieved from: <http://www.prisonexp.org/>

Appendix 1: Games Discussed

Arcane Gallery of Gadgetry

Purpose: Educate students in the history of the
US Patent Office

Site: <http://www.arcanegalleryofgadgetry.org/>

ARGOSI

Purpose: orientation for university students

Site: <http://argosi.playthinklearn.net/>

Art of the Heist

Purpose: Audi marketing campaign

Site: <http://campfirenyc.com/work/audi-art-of-the-heist>

The Beast

Purpose: A.I. movie marketing

Site: <http://www.seanstewart.org/beast/intro/>

Come Out and Play

Purpose: festival series to encourage public games

Site: <http://www.comeoutandplay.org/>

Conspiracy for Good

Purpose: Encourage involvement with nonprofit causes

Site: <http://conspiracyforgood.com/>

Coral Cross

Purpose: CDC game for pandemic preparedness |
and response

Site: http://www.argn.com/2009/04/coral_cross_pandemic_preparedness_from_the_hawaii_department_of_health/

Edoc Laundry

Purpose: Entertainment; monetization through
integrated t-shirt sales

Site: <http://sml-design.com/edoc-laundry/>

Foresight Engine

Purpose: Institute for the Future framework for foresight scenario creation

Site: <http://www.iftf.org/what-we-do/foresight-tools/collaborative-forecasting-games/foresight-engine/>

FutureCoast

Purpose: Explore environmental foresight scenarios at a personal level; motivate environmental activism

Site: <http://futurecoast.org>

Geocaching:

Purpose: Entertainment; monetization through accessory sales

Site: <http://www.geocaching.com/>

Ghosts of a Chance

Purpose: Smithsonian American Art Museum outreach

Site: <http://www.ghostsofchance.com/>

Heroes 360/Heroes Evolutions

Purpose: Heroes TV show marketing and extended universe

Site: <http://www.nbc.com/heroes/evolutions/>

I Love Bees

Purpose: Halo video game marketing

Site: <http://www.ilovebees.com/>

Ingress

Purpose: Entertainment, beta test for Niantic platform,

Google data collection

Site: <http://www.ingress.com/>

Insectopia

Purpose: entertainment

Site: http://iperg.sics.se/iperg_games7.php

Letterboxing

Purpose: entertainment

Site: <http://www.letterboxing.org/>

The Lost Experience

Purpose: Lost TV show marketing and expanded universe

Site: <http://thelostexperienceclues.blogspot.com/>

Meigeist

Purpose: entertainment

Site: <http://www.dshed.net/dshed/meigeist>

Maker Cities

Purpose: foresight planning

Site: <http://www.iftf.org/our-work/people-technology/technology-horizons/maker-cities/>

Metacortechs

Purpose: entertainment (independent game based on The Matrix movies, not created as marketing)

Site: <http://www.metacortechs.com/>

MMOWGLI (Navy)

Purpose: foresight planning for U.S. Navy

Site: <https://portal.mmowgli.nps.edu/game-wiki/-/wiki/PlayerResources/About%20MMOWGLI>

Momentum

Purpose: entertainment

Site: <http://www.digra.org/wp-content/uploads/digital-library/07313.58398.pdf>

The Optimist

Purpose: marketing for Disney

Site: <http://optimist.disney.com/>

Perplex City

Purpose: entertainment, monetization through sales of puzzle cards

Site: <http://www.perplexcity.com/>

Reality Ends Here

Purpose: student orientation and creation of artistic collaborations

Site: <http://reality.usc.edu/about/>

RevQuest

Purpose: education and tour guidance at Colonial Williamsburg

Site: <http://www.colonialwilliamsburg.com/do/special-events/revquest/>

Second Life

Purpose: social interaction and creative expression; also used for conferences, etc.

Site: <http://secondlife.com>

Survive DC

Purpose: entertainment

Site: <http://www.survivedc.com/>

Tower of Babel

Purpose: second language education

Site: <http://arg.uws.ac.uk/>

The Walk

Purpose: encourage exercise

Site: <https://www.thewalkgame.com/>

World of Warcraft

Purpose: entertainment

Site: <http://us.battle.net/wow>

World Without Oil

Purpose: Create and explore foresight scenarios around peak oil

Site: <http://worldwithoutoil.org/>

Zombies Run

Purpose: encourage exercise

Site: <https://www.zombiesrungame.com/>

Zooniverse

Purpose: Engage players in crowdsourced analysis of scientific data

Site: <https://www.zooniverse.org/>

Appendix 2: Recommended Readings

The following resources, out of all those cited here, seem most relevant and useful to the creation of ARG-based research:

Books

Pervasive Games: Theory and Design, by Marcus Montola, Jaakko Stenros, and Annika Waern.

Impressive overview of methods, risks, and case studies.

A Creators' Guide to Transmedia Storytelling: How to Captivate and Engage Audiences Across Multiple Platforms, by Andrea Phillips.

Aimed at creators of both games and passive entertainment, but includes detailed and current discussion of best practices.

Articles

Spies like me: My response to IARPA's RFI UAREHERE, by Carrie Cutforth-Young. Available at: <http://queenspade.com/my-open-response-to-iarpas-rfi-uarehere/>

Detailed, blunt discussion of ethical and practical concerns around ARG-based research.

Fischer, J. E., Flintham, M., Price, D., Goulding, J., Pantidi, N., & Rodden, T. (2012). Serious mixed reality games. In *Mixed Reality games: Workshop at the 2012 ACM Conference on Computer Supported Cooperative Work*.

One of the few papers to explicitly discuss the potential of ARG-based research

Grant, A. M., & Wall, T. D. (2009). The neglected science and art of quasi-experimentation: Why-to, when-to, and how-to advice for organizational researchers. *Organizational Research methods*, 12, 653-686.

Overview of the strengths and risks of mixed methods.

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61-135.

Seminal paper on participant diversity issues.

International Game Developers Association. (2006). Alternate reality games white paper.

Excellent introduction to what ARGs are and what they can do.

Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science*, 7, 109-117.

Good source of ideas for research topics—discusses areas that do and don't replicate well in the lab.

Appendix 3: What a Research ARG Might Look Like

This ARG is built, as many are, around the intrigue/thriller and science fiction genres. Participants are recruited as “agents” for one of two covert organizations, each trying to find and make use of a hidden alien technology. Play takes place within a limited geographic area (e.g., a particular city) over the course of a month. Players use experimenter-provided devices (GPS or phone) that record interactions. They also use online blogs and social media sites that are intended to be faction-specific, but can be hacked into by the opposing faction.

- Incentives: Players are paid for playing (or given in-game currency that has value during play) and can gain or lose depending on particular aspects of performance.
- Goals: Each agent is assigned to seek out a particular piece of information about the alien technology, which is held by another player. They also have information of their own, certain other players to whom they are incentivized to give it, and others they are incentivized to deceive about it.
- Measures: Players complete a “recruitment questionnaire” at the start of the game that collects baseline attitude, personality, and mood measures. “In the field” their organizations require them to wear chest straps to measure heart rate and breathing. They are periodically either asked to take other readings on themselves, or pulled aside by “superior officers” to get more extensive readings. Data about behavior and trust-related choices are collected through the game device and dedicated websites.
- Variables: competence, benevolence, and integrity as components of trust
 - Competence: Players have to complete a medium-difficulty puzzle to get their own information (justified by the nar-

rative as a test put in place by the aliens). The better they do on the puzzle, the better and more accurate their information. The puzzle is normed in advance to get the desired distribution.

- Benevolence: Teams of players have competing goals. Fooling opposing players gets you closer to your goal, while pushing them farther away from theirs. Teams also engage in a shared task together at the beginning of the game, in competition with each other, to increase players' liking for their own side and decrease benevolence towards the other.
- Integrity: Some players are incentivized to betray their own side, either by being given a conflicting individual goal, or through bribery by game-runners taking on in-game roles. This is the hardest variable to put in place, but experimenters could continue making offers until they have a good percentage of traitors. For comparison or as an alternative, they could simply inform some people that they're playing traitors—this is easier, but probably not as psychologically valid as players actually choosing to give up their integrity within the game.

The game could test hypotheses regarding what factors affect different components of trust, and the timelines over which they are evaluated. For example:

- Participants attempting to evaluate trustworthiness might be aware that some agents have low competence from the start, but come to weight integrity more heavily over the course of the game.
- Different stages of team formation (e.g., forming versus storming), or stressors such as insufficient resources, may affect the difference in benevolence towards one's own faction and the opposition.

- Specific physiological variables, picked up by the portable sensors described above, may predict trustworthy behavior or accurate judgments of trust.

Such a game includes considerable room for flexibility. It could be replicated at different sites, and with different populations based either on location or deliberate recruiting differences. It could be extended to look at longer time frames, or to include additional variables and hypotheses related to other types of human interaction. In supporting the creation of fully developed, high-context communities, research ARGs may provide a broad canvas for exploring a wide range of behavioral, social, and neurophysiological questions.

Acknowledgments

Books, like ARGs, are ultimately a collective endeavor. In order to put this one together, I had to track down clues, gather information across media, and interact with a wonderful variety of fascinating characters. Any errors are mine, and shouldn't reflect on the rich insights and experiences of the many people who helped guide my way. Special kudos are due to a few:

Adam Russell was this project's puppetmaster: his ideas and questions started me on my journey of discovery, and his encouragement and feedback were instrumental throughout. Alexis Jeannotte carried out our first explorations of how alternate reality games might fit with behavioral research, and her notes and advice kept me on the right track.

Several people generously shared their experiences with ARGs and ARG-like activities. Mark Ruppel, of the National Endowment for the Humanities, sat down with me early on to discuss his experiences with serious games of all sorts, and to brainstorm about what ARGs might have to contribute. Rita Bush also shared her insights into federal use of serious games. Lisa Fischer, designer of Colonial Williamsburg's RevQuest game, gave me a wonderful run-down of how ARG design and revision play out on the ground. Georgina Goodlander took a couple of precious hours—a week before moving across the country!—to share lessons learned from the Smithsonian's Ghosts of a Chance and Vanished games. Matthew Largent helped me better understand the Navy's experiences adapting the Foresight Engine model to their own internal problem solving efforts.

Others gave me a more direct window into on-the-ground ARG design and play. Kari Kraus and Elizabeth Bonsignore, along with their collaborators and students at University of Maryland and Brigham Young University, invited me to a design session for their in-progress educational

ARG. Watching them create a working rabbit hole was itself educational, and I look forward to hearing more about the tardigrades. David Tomblin's DC instantiation of the Arizona State University Futurescape Tours helped me see how media and live experience can combine to create larger societal insights. Ken Eklund sent me on a mission to hide a voicemail from the future. And Roc Myers, in addition to providing insightful feedback on the manuscript, helped an unenlightened newbie overcome her resistance to figuring out the Ingress interface.

Mark Deloura, Dan Laughlin, Anna Muldoon, and the other members of Federal Games Working Group provided ongoing suggestions and feedback on the general topic of research ARGs, as well as mockery for my slow progress and choice of faction in Ingress. Lea Shanley, Jay Benforado, Rebecca French, and the other members of the Federal Community of Practice on Crowdsourcing and Citizen Science also shared suggestions and feedback. Michael Morefield of Arctan Inc. gave feedback and suggestions, and suggested several thought-provoking ways to frame ARG-related activities.

Along with in-depth feedback and discussion of her own design experiences, Brooke Thompson put me in touch with Christy Dena, who in turn connected me with ETC Press. Many thanks to both of them for starting this book on the road to publication.

About the Author

Ruthanna Gordon received her PhD in Cognitive Psychology from Stony Brook University in 2003. Her research focused on memory, decision-making, and narrative processing—and how they affect human interaction in everyday life. The limitations on lab research described in the book are ones that frustrated her in practice, and many of her studies involved gamification and storytelling as strategies to improve real world applicability.

In 2011 she switched from academia to policy work, and swiftly became involved in the new movement to integrate serious games in the government's tool box. As an American Association for the Advancement of Science policy fellow with the U.S. Environmental Protection Agency, she worked on the creation of a mixed reality game—intended to include both online components and live missions—to help households learn about and reduce stormwater pollution. She currently works for AVIAN LLC, where she supports innovative federal research in the social sciences.

She's also a lifelong gamer, with experiences ranging from Chess to transmedia and pervasive games. She participates in tabletop roleplaying (mostly *White Wolf* and *Dungeons & Dragons*) as well as occasional LARPs. She's Enlightened, although not very.

