Risks and Rewards of Rapid Play Games

This chapter describes *rapid play digital games* as a method for emergency management education. It discusses the theoretical benefits, common pitfalls, practical techniques to avoid those pitfalls, and some templates to kick-start design. Applying these techniques and templates will help ensure that an instructional game for professionals is effective, targeted, and affordable.

A *rapid play* game is one that is playable in less than 30 minutes, allowing a participant to complete multiple independent scenarios in a 1–2 hour sitting (R.M. Seater, 2018) (Seater, 2019) (B. Vogt, 2015). This format supports highly iterative play with several key benefits:

- **Immediate feedback** and support of a try–fail–retry pattern of content exploration.
- **Repeated exposure** to improve retention and build mental models of the tradeoff space.
- **Training on rare, costly, and dangerous events** that offer inadequate or unsafe opportunities for field experience but are complex enough to need multiple iterations to build intuition and experience.
- **Exposure to a wide variety of scenarios**, putting similar decision points in different contexts, thereby requiring different responses and pursuit of different priorities.
- **Collect quantitative performance data** to help instructors assess the needs of individual students or of the broader curriculum.
- **Increased engagement** by embedding abstract lessons into concrete scenarios, following the lessons from naturalistic decision making (Klein, 1998).

Many of these benefits are shared by other forms of education, gaming, and simulation. Rapid play games are defined by their short playtime and flexibility to express a different scenario on each play. The category is broad, and it includes some immersive simulators (e.g., flight simulators) and some tabletop games. This chapter focuses on rapid play *digital* games, although many of the techniques described also apply to non-digital games and non-rapid games. Rapid play games bridge methods from the mod-sim and game-based instructional communities (Page, 2019).

The ability to see multiple scenarios in a single sitting is the main reason to use a rapid game over a more immersive and detailed (and longer) game. Rapid play games are valuable when seeking to build flexible mental models for decision making in uncertain, ambiguous, and dynamic situations. A longer game (or non-game method) might be more appropriate when the learning objectives focus on ingraining procedure, training on the use of particular equipment, building team cohesion, or honing raw technical skills.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

The potential benefits of rapid play games do not substitute for good design and good requirements. The design of a rapid play game, like other game-based learning methods, is riddled with pitfalls. Four such pitfalls stand out as the most common and problematic.

- **Quiz Pitfall.** A digital game can devolve into a series of quiz questions, testing a student's factual knowledge rather than serving as an interactive system to help them understand strategic implications, tradeoffs, and context. Quizzes are not bad tools, but games are an inefficient way to administer them.
- **Bloat Pitfall.** It can be tempting to build a simulation of all facets of the environment in great detail, thereby burying the key lessons of the game, raising the cost of development of the game, and increasing the time commitment required by the student to complete the material. Detailed models are valuable for immersive training simulators, but they are often unnecessary for getting at instructional goals related to tactical and strategic decision making.
- **Irrelevant Mechanics Pitfall.** Gameplay mechanics from successful entertainment games and genres should not be included without considering whether they are necessary, whether they support the core learning objective, or assuming that imitating superficial properties of successful games will lead to effective outcomes. Reusing an entertainment game template can save time and engage players, but it can also be a distraction to players and developers.
- **Silver Bullet Pitfall.** When designing a serious game, one should critically assess which material should be conveyed via a digital game vs. other instructional mediums – a short film, digital quiz, case study or historical anecdote, field exercise, round table discussion, tabletop game, presentation by a survivor, or traditional classroom presentation by an expert. Like any tool, rapid play games are not right for every job; they fill an important niche that complements other methods.

This chapter describes eight techniques to avoid those pitfalls and maximize the potential gains of using rapid play instructional games. It also describes five game templates—design patterns that embody the described techniques and serve as a starting point for building an effective rapid play game.

These concepts will be illustrated on a running example of a fictional game called "*Quake Responder*" about allocating scarce resources in the aftermath of a major earthquake. Variations of this game will be introduced throughout the chapter to illustrate the techniques and templates. Techniques 1–4 will be applied to a strategic version of the game about allocating scarce resources between competing objectives. Techniques 5–8 will consider a more tactical variant involving an added radiological threat. The templates will each consider a different flavor of the scenario to illustrate the style of game best suited for each.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

Technique 1: Focus on Dilemmas

A practical means of avoiding common pitfalls is to focus on key *dilemmas* relevant to the learning objectives, rather than just diving into modeling the domain from the bottom up. A dilemma is a decision or set of related decisions that has no correct answers *in isolation*, but which directly affects critical mission outcomes and thus does have a best answer *in context*. A decision might be a dilemma because there are high consequences for getting it wrong, because time and information are insufficient to predict all consequences, or because the best decision is highly dependent on shifting priorities and context (Adams, 2013) (Burgun, Game Design Theory, 2012). The following decisions are dilemmas:

- Whether to evacuate a population center when a hurricane forecast and projected path are still uncertain
- Whether to proactively deploy national stockpiles in response to a potential infectious outbreak prior to confirmation of the event
- Whether to allocate limited resources towards immediate lifesaving or evacuation preparation
- Whether to push back against a collaborating organization that is behaving sub-optimally or support their approach to maintain cohesion and trust

Technical knowledge of the domain will help frame and guide the decision but not render the choice obvious. If a decision is obvious to an expert, it is not a dilemma. The following decisions are not dilemmas:

- Which hurricane forecast tool provides the most accurate forecasts at a given time horizon
- How to select an optimal evacuation route or how to optimize a distribution network
- What doctrine says to prioritize in a given situation
- The legal and procedural steps to establish collaborations with other organizations

These are fine lessons to teach, and they can be conveyed in a rapid play game, but they do not justify the use of a rapid play game.

A rapid play game is well suited to exploring dilemmas, since it allows the student to experience the same choice in different contexts and understand the emergent implications and tradeoffs. One of the benefits of a rapid play game is that, by the nature of its brevity, a student can experience different scenarios in a single sitting. Scenario diversity helps prevent overlearning or overgeneralizing lessons from one example by forcing the student to confront the same decision in different contexts with different response patterns.

For the Quake Responder sample game, a key learning objective might be to correctly balance saving lives in the short-term with rebuilding the economy in the long-term. A dilemma in the game might arise at a decision point when the player must either divert resources to restoring hospitals (immediate benefit) or to rebuilding casinos (revenue to fund future recovery efforts). The best decision will vary

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

depending on context—the scale of the disaster, the time that has passed since the disaster, the region's economic dependence on casino tourism, and the availability of substitute medical care in nearby regions. On different plays of the game, the player faces different combinations of those factors (perhaps from a random scenario generator or perhaps from a large pool of hand-crafted scenarios), changing the ideal allocation of resources. Students cannot score well over a series of games by always taking the same action; in-game success requires building a richer mental model of the contextual factors to focus on and the tradeoffs they entail.

A focus on dilemmas helps address the aforementioned pitfalls.

- **Quiz.** A focus on dilemmas keeps the game focused on topics where interactivity and repetition are beneficial and away from quizzes of technical facts, doctrine, or written policy. In Quake Responder, technical knowledge of how to estimate the cost of rebuilding a hospital or the long-term health impacts of a sluggish economy are helpful, but those pieces of factual knowledge must be applied to a context.
- **Bloat.** Use the list of key dilemmas as a criteria for eliminating content bloat; any detail or dynamic not relating to one of the identified dilemmas can be omitted. In Quake Responder, there is no need for a map of the city or a model of evacuation routes – as these are details that do not inform or enrich the dilemma of resource allocation.
- **Irrelevant Mechanics.** Create or borrow gameplay mechanics that serve and emphasize the dilemmas rather than obscuring or distracting from them. A first-person three-dimensional city view in Quake Responder would be flashy but irrelevant to exposing the target dilemma. In a game about predicting flood zones, a three-dimensional view might be critical.
- **Silver Bullet.** If the learning objectives do not overlap with the identified dilemmas, a rapid play game is probably not the best instructional tool. If Quake Responder's main learning objective is to build empathy with survivors or drill cost estimation, a game will be less impactful than a short film or series of test questions.

Technique 2: Include Multiple Scoring Criteria

Providing a student with a single numerical score after a playthrough is a powerful part of what makes games different from (and sometimes more useful than) other interactive experiences. When a player experiments with different approaches, it is the score that helps them recognize successful strategies. The score helps the player build a mental model of what actions are good or bad in a given context by summarizing the net effect of complex interactions. Feedback is critical to learning, but providing clear feedback can be difficult when modeling complex environments (S.A. Yourstone, 2008).

Emergency response often features competing measures of success. The idea of providing a single numerical score can be at odds with teaching students about the complexities and competing priorities

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

of a real event. Typically, this problem is reconciled by computing a final score as a weighted sum of several component scores.

For Quake Responder, the two competing scoring objectives might be quantified and combined as follows:

- H = Health Subscore = percent of patients who receive adequate treatment
- E = Economic Subscore = average standard of living 10 years later
- Overall Score = 10 H + 0.75 E

The game would have different incentives if the overall score were given as *1.75 H + 5 E*, even though the same two factors would be emphasized. The two weights (10 and 0.75, or 1.75 and 5) are 'magic numbers' that the game designer uses to quantify the relative importance of the two factors and to scale them to be comparable. They reflect the designer's judgment, hopefully properly informed by the experts, of the relative importance of the two factors.

It is often a mistake is to make those weights constant—to always have the same score equation for all scenarios and all game iterations. Doing so undermines the notion that priorities change with context and might build bad habits in the student. Consider the following alternatives to a fixed scoring formula:

- **Vary Weights Based on Scenario.** Use a different set of weights for each scenario, but do not reveal those weights to the student until the scenario is over. In a scenario set two days after a large earthquake, weigh the health subscore higher. In a scenario set two months after a moderate earthquake, weigh the economic subscore higher. This approach forces the player to infer the weights from the scenario and requires that they understand the relation between context and priorities. It is a good approach when the experts creating the scenarios agree on what the proper tradeoffs should be in different situations.
- **Randomly Vary Weights.** Randomly vary the weights on successive plays of the same scenario and reveal those weights to the player prior to each scenario. In one scenario, the player might be told to weigh health higher, whereas in the next game, they are told to weigh economics higher for the exact same situation. This approach forces the player to adapt their strategy based on shifting priorities. It is a good approach when the player's real-world role does not have the authority to determine priorities but does need to adapt to different statements of priority from superiors.
- **Report Independent Subscores.** Report each subscore back to the player without combining them into a single number. A player would get a health score and an economic score, with *no* statement of their relative importance. This approach is suitable when the game does not (or cannot) capture all of the relevant factors that would determine the relative weight. Those might be factors are unknown, unknowable, or too complex or informal to encode in an

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

abstracted simulation. It is suitable when you want students to understand *how* their decisions affect the two subscores but do not want to take a stand on how those subscores *should* be balanced.

If the game is modeling a response to a routine house fire (with expert agreement and strong historical precedent), it is best to vary weights based on the scenario. If the game is modeling a fire chief's response to a large-scale earthquake (an event much larger than the player's authority), randomly varied weights are appropriate. If the game is modeling a multi-organization response to an improvised nuclear detonation (a rare event that does not have agreed upon priorities), independent subscores are appropriate.

When identifying relevant scoring subcomponents, consider the following types of competing objectives:

- **Organizational Values.** The relevant authority or organization often has stated values that are at odds with each other. In Quake Responder, the player's organization values both the provision of health services and the restoration of economic prosperity, making those natural subscores to put in opposition.
- **Time Horizons.** Even with a single clear objective, there can be tradeoffs between short- and long-term outcomes. In Quake Responder, the player might only be pursuing lifesaving but be forced to trade off short-term (search and rescue), mid-term (medical stabilization), and longer-term (evacuation to major hospitals) aspects of lifesaving.
- **Multiple Authorities.** Each organization might have a single goal on a single timeframe that differs from other organizations and authorities. A game of Quake Responder might be scored against the imperfectly aligned priorities of FEMA, local political leaders, and collaborating non-government organizations (NGOs).
- **Moral Obligations.** There may be moral imperatives apart from stated organizational goals. A player in Quake Responder might have to trade off saving lives and reducing suffering. A player might have a decision point of how much medical care to divert to victims who are unlikely to survive even with treatment, but who are suffering in the short-term. A given organization might have no official policy on how to reconcile those goals but still want employees to make informed tradeoffs in the moment. A rapid play game can help a player to understand such tradeoffs by seeing them in a variety of contexts, even when there is no agreed upon 'best response' by experts.

Pitfalls

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

- **Quiz.** Varied scoring formulae are great for building mental models but can complicate objective measures of technical skill. This technique helps avoid the quiz pitfall by maintaining a focus on contextual tradeoffs rather than context-independent truth.
- **Bloat.** Be wary of including too many scoring criteria. The bloat pitfall applies to scoring detail just as much as modeling detail. Choose the scoring criteria that are relevant to the target dilemmas and learning objectives, and avoid the temptation to include more just for the sake of realism or fidelity. Often, a game only needs two competing criteria to get a message across and create an appropriate dilemma.
- **Irrelevant Mechanics.** Entertainment games almost always have constant scoring weights, since they are heavily focus on competition and fair comparison. However, narrative and sandbox games often have more ambiguous and multi-faceted success criteria—they generally show players the long-term consequences of their actions but make no quantitative judgment on whether that outcome was optimal. When drawing inspiration from entertainment games, think carefully of whether the scoring system should be preserved (for familiarity) or overhauled (to match the learning objectives).

Technique 3: Procedurally Generate Scenarios

"Don't design the thing; design the process that produces the thing." –Neil Gershenfield

One drawback of professional military wargames is inflexibility, and that inflexibility often persists into non-military games that intellectually inherit from those roots. Many wargames are designed to only support a single scenario of interest—a particular city geography, a particular sequence of surprises, and/or a particular pool of response assets. Adding one additional scenario can be nearly as costly as creating the entire game in the first place. Such games can model one scenario in great detail, but they are limited in breadth.

A rapid play game's utility lies in allowing a student to play many different scenarios in rapid succession. The designer of such a game must therefore either commit to handcrafting many scenarios or use automatic procedural generation of scenarios.

Even if each student only experiences a few scenarios, there can still be value in procedural generation to reduce cheating and sequencing effects. Pilots often tell stories of arriving at a flight simulator test and being told by the prior student "be ready for a squall 10 minutes in". Unsurprisingly, later students score better than earlier students in a manner unrelated to actual ability. Rather than fight the culture of cooperation among the students, make that kind of borderline cheating impossible by having every scenario be slightly different. If the best advice a student can give the next person in line is "be ready for squalls at some point" or "be ready for some kind of disruption 10 minutes in" or even "be ready for

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

some kind of disruption at some point", the sequencing effect is reduced and the targeted skill is more directly tested.

Think of the procedural generation logic as a pipeline. Start with the key contextual factors you want students to learn to look for, then add semi-random derived variables to enrich the scenario, and finally compute scoring weights appropriate for that situation. A Quake Responder scenario might be procedurally generated as follows:

1) **Identify key contextual factors**
    a. Scale of the disaster: 100k, 500k, 1 million, or 3 million people affected by the event
    b. Time since the disaster: 1 day, 1 week, 1 month, or 6 months since the event
    c. Economic dependence on casinos: 5%, 25%, 50%, or 75% of tax revenue comes from casinos
    d. Nearby substitute medical care: 0%, 25%, 50%, or 75% of population can be served by nearby facilities
2) **Add derived variables**
    a. The overall level of national demand for casinos is randomized independently of other factors.
    b. The percentage of civilians needing medical care is computed based on the scale of the disaster, the time since the disaster, and a random factor.
    c. The level of tax revenue from non-casino sources is computed based on the scale of the disaster, economic ties to nearby regions, and a random factor.
3) **Compute scoring weights**
    a. A player's health subscore is based on the number of people affected by the disaster, the percentage of people needing medical care, the capacity of local facilities (after the player restores some of them), and the capacity of nearby facilities.
    b. A player's economic subscore is based on the number of people affected by the disaster, the economic dependence on casinos, the overall national demand for casinos, and the level of tax revenue from non-casino sources.
    c. The weight between the health and economic subscores is computed based on time since the event—if more time has passed, more weight is placed on economics. Alternatively, randomly select the health and economic weights, or report the two scores independently (as per Technique 2).

To generate a particular scenario from the above template, randomly select the items from Step 1 with uniform probability, then semi-randomly compute derived variables as described in Step 2, then deterministically compute scores and weights in Step 3.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

Note that the procedural generation algorithm does *not* include parameters for the layout of the city, its vulnerability to flooding, or the economic model of nearby regions. Those are interesting factors relevant to other dilemmas, but they would distract from the dilemma targeted by this game. They are not even included as derived variables. The decision to make a given factors a key variable, a derived variable, or an omitted variable should fall to the domain experts and the team who provided the learning objectives—the job of the game designer is to reflect those priorities in the game and the scenarios.

If the game requires very detailed scenarios, consider a Bayes Net to represent the interconnected conditional probabilities of key and derived factors (E. Grois, 1998). For simple scenario structures, such frameworks are excessive. For complex scenarios, they are more maintainable and can simplify the arithmetic of generating internally consistent scenarios.

Pitfalls

- **Quiz.** Procedural generation forces the game to generate novel situations, not just run students through canned scenarios with rigged solutions. Hand-crafted scenarios can feel like a quiz with one right answer. The unfairness of random scenarios can be more memorable than playing scenarios where a perfect score is always possible (E. A. Kensinger, 2003).
- **Bloat.** Keep the list of key factors short, or else procedural generation will become overly complex or produce non-credible scenario combinations. Any factor that isn't critical to the learning objectives or doesn't relate to your target dilemmas should be relegated to a derived variable, held constant for all scenarios, or omitted entirely. The goal is not to build a realistic model of the world—the goal is to highlight one key relationship within a complex world.
- **Irrelevant Mechanics.** Procedural generation is a common mechanic in entertainment games, especially in video games. However, competitive games often focus on consistency and fairness over variety and flexibility (S. Garozzo, 2015). The type of level variety provided by an entertainment game might be 10 variations of one scenario not 10 scenarios that make the student react flexibly. When using an entertainment game for inspiration, think carefully about what type of scenario variety will serve the learning objectives.
- **Silver Bullet**. If not done well, procedural generation can produce unbelievable scenarios that harm education and credibility. If there is not time to do procedural generation carefully, consider creating a pool of highly varied hand-crafted scenarios as a quick alternative.

Technique 4: Include Quiet Scenarios

One reason why emergency response is difficult in the real world is incorrect, incomplete, and untimely information about what is really going on. It is easy to lose those effects in a game, thereby losing instructional value. Players may apply 'student reasoning' to guess correct answers without following

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

the correct deductive path to get there, simply because they know that the game is artificial (Chekhov's Gun, 2020) (The Law of Conservation of Detail, n.d.). Any aspect of a scenario a student can anticipate for artificial reasons is a lesson they did not learn about the real world.

Rapid play game offers an opportunity to undermine such distortions with randomly interspersed *quiet scenarios*. Since a student is expected to play multiple scenarios in sequence, it is not necessary for every scenario to have an event. Some scenarios can be quiet scenarios where no event happens, but misinformation might mistakenly lead a player to waste resources to no effect. The majority of scenarios can still test a student's ability to handle an actual threat, while the occasional non-event scenario preserves elements of uncertainty and surprise.

Suppose that some Quake Responder scenarios are procedurally generated to have no medical shortfall and no economic dependence on tourism. In such cases, neither sector requires resources. Even if only one of five scenarios have zero demand, a player will have to consider that possibility in every scenario. They cannot ever just allocate resources to the greatest need, and they are also required to assess whether the need is present. The plays in which there is no need are themselves less educational, but their existence, even in small numbers, improves the instructional value and realism of the rest.

Pitfalls

- **Bloat**. The game doesn't need to model out the full set of actions a player does once they identify a scenario as being quiet. The game can end with the player recognizing the absence of a threat. The quiet scenario does its job even if it offers no meaningful gameplay or feedback loop.
- **Irrelevant Mechanics**. Quiet scenarios are rare in entertainment games, which often seek to maintain a constant state of flow for the player (Csikszentmihalyi, 1990). Emergency management is fundamentally about disruptions to a quiet norm, not maintaining an efficient status quo.
- **Silver bullet.** This technique is specific to rapid play games. If running a longer game with only time for one session, the surprise benefit of having no threat present is unlikely to be worth spoiling the only play of the game.

Technique 5: Stretch the Reality of Rarity, Transparency, Authority, and Restrictions

"It's not wise to violate rules until you know how to observe them." –T. S. Eliot

"By all means break the rules, and break them beautifully, deliberately and well." –Robert Bringhurst

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

By default, instructional games should aim to be realistic. In key aspects, however, games may intentionally deviate from realism in service of the learning objectives. When building a rapid play game, there are four main cases where realism sometimes might take a back seat.

- **Frequency of Rare Events.** Rare events should made unrealistically common in game scenarios so that they actually occur. As per Technique 4, there still need to be *some* quiet scenarios but not a realistic number of them. An overly realistic game about large-scale earthquakes would almost never present the student with an earthquake!
- **Level of Information Transparency.** Information that is unavailable, noisy, or deceptive in the real world should often be (initially) made available, clean, and accurate in an instructional game. Players will build mental models better if they see accurate and immediate consequences of their actions. It is often helpful to have two game modes: an 'easy' mode with exaggerated information transparency used to build basic understanding, and a 'hard' mode with realistic levels of information transparency to test that mental model and prevent players from becoming reliant on the transparency.
- **Breadth of Authority.** Serious games should clearly specify the player's role. However, not all limitations of that role should be enforced. Players should sometimes be permitted to take actions just outside of their role's true authority in order to build understanding of how different roles can effectively coordinate. To avoid teaching the wrong lessons, force players to explicitly shift in-game roles in order to access actions requiring different real-world authorities—make them aware of where the boundaries are, but let them move between the boundaries to understand what a smoothly coordinated response can look like.
- **Doctrine and Policy**. Even a game that is intended to teach standing doctrine or policy should sometimes allow a player to deviate from such restrictions. Doing so helps the player build a mental model of the dynamics of the situation and appreciate *why* the doctrine and policy are in place. To avoid teaching the wrong lessons, note to a player during scoring (or as they play) when they deviate from doctrine or policy.

All of these deviations from reality are in service of increasing feedback to the student and providing that feedback on a shorter timescale, with all the benefits that come with rapid feedback (Hattie, 2012) (K. L. Kettle, 2010).

Consider how those concepts might map to Quake Responder. Augment the game to include a potential radiological aspect—in some scenarios, players are responding to a site that has a radiological threat as a secondary consequence of the original earthquake. The player is being trained how to handle radiological events when specialists are not available and when information about the nature of the threat is incomplete.

- **Frequency of Rare Events.** Actual radiological events are rare, and it is even rarer that a specialist team will not be available. The game generates scenarios that have a radiological

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

element 75% of the time and no radiological component (but still some misleading indicators) 25% of the time. Players mostly gain experience managing the complications added by a radiological threat, but they cannot take for granted that every incident will involve radiation. Their performance will suffer if they ignore the radiological threat, but they will also fall short if they assume that there is always a threat and act overly cautious in all scenarios.

- **Level of Information Transparency.** Radiation can cause harm without immediately visible symptoms, and in a real incident, a responder may not have proper sensing equipment. In the game, players first play on easy mode with a perfect dosimeter giving the current rate of exposure, total accumulated exposure, and safe limits of exposure. As a player progresses to harder difficulties, those pieces of information are taken away or degraded. Players build intuition for how different actions lead to different exposure, then they are required to operate without the artificial feedback.
- **Breadth of Authority.** Actual first responders performing search and rescue do not typically have authority to decide their work shift durations. In the game, however, the player is allowed to freely switch between controlling the tactical response and controlling staff time allocation. By having control over two related roles, they build a mental model of how to balance lifesaving efficiency with managing exposure levels of the responders. That mental model also helps to build trust and empathy between the two roles, so that each understands the limits and pressures the other faces that might otherwise cause tension or conflict within the team.
- **Doctrine and Policy**. Under ideal circumstances, standing policy will dictate what a safe radiation exposure level is, and someone building a shift schedule would not deviate from such guidance. In the game, the player is allowed to deviate, and the game models the consequences of doing so (both to the responders and to the number of lives saved). The player is allowed to take risks with their team when the possible gains are high or protect the response team when likely gains are low, even if policy would normally forbid such deviations. In doing so, they learn the tradeoffs they may face in non-ideal circumstances and build an understanding of why the policy sets the bounds that it does. In a real incident, a responder needs to trust the policies they follow and know when the situation warrants deviating from them. A game that never lets the player cross policy lines will never build intuition for why those lines are present.

Pitfalls

- **Quiz**. Because of the risk of misinformation, it can be sensible to pair a game with a quiz on the topics stretched in the game to confirm that the students understand where the game intentionally deviated from reality in service of the learning objectives.
- **Bloat**. Do not stretch reality in all these ways as once. Pick the subset that targets the learning objectives and selected key dilemmas, and keep all other aspects of the game as realistic as

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

possible. Every deviation from reality is one more piece of information that will take up time when explaining the game or discussing it after play.

- **Irrelevant Mechanics**. When borrowing structures from entertainment games, be wary of where this principle has been taken too far. At a minimum, make it clear to the player where the game deviates from reality when introducing it or discussing it afterward.
- **Silver Bullet**. This method is powerful and dangerous, and it should be used judiciously and with close oversight from domain experts. Make sure that the deviations from reality are not counter to the learning objectives, and that the loss of reality is justified by a commensurate gain to those objectives. Accompany the game with a description of elements that were intentionally altered for the sake of education, so that an instructor using the tool can properly frame and introduce it.

Technique 6: Analyze Degenerate Strategies

Much of the literature about serious games (and education in general) focuses on the question of transference—whether lessons learned in a simulated environment translate to a real environment (T.M. Connolly, 2012) (Zhonggen, 2019) (P.M. Kato, 2008) (Conde-Pumpido, 2017) (R. Seater, 2016). An equally important question is misinformation—whether the game is teaching the *right* lessons. An instructional method that teaches bad habits can be more harmful than one that fails to teach anything at all, and rapid play games are no exception. One of the powers of a game is the feedback provided by getting a score after every iteration. However, rapid feedback can backfire if players learn to “game the system” or exploit the game’s abstractions and simplifications in unrealistic ways.

A rapid play game should teach the right lessons even with players who are motivated only by score. Do not ask students to “play in the spirit of the game” or “just behave realistically” (which can be reasonable guidance in other types of instructional games). Even if players are highly motivated to learn, telling them to “play in the spirit of the game” can undermine the value of using a game—it implicitly tells them to not explore boundaries, not try new strategies, and not trust what they learn from the game. Telling students to “just behave realistically” presumes they know the lesson the game is being used to teach. The iterative nature of a rapid play game makes it more likely that a student will discover a degenerate strategy, so that particular strength of rapid games is a liability if the game is not calibrated carefully.

The main property of rapid play games that creates the risk of exploitation is the simplification and abstraction necessary to manage duration and complexity. The game must simplify the real world to fit into a time slot, focus on a particular lesson, and provide clear feedback. Those simplifications can create exploitable edge cases counter to the intended lesson. Even simple games can have subtle

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

exploits that designers miss. Fixing such problems is often easy once they are identified, but catching them in the first place is challenging.

Fortunately, the very property of games that amplifies this risk—simplification of the real world—offers a mitigation. A game that has been abstracted and simplified down to a focused reaction model, short list of valid actions, and prescribed set of possible feedback is also amenable to analysis in the form of mathematical computation, human testing, and/or computer-aided optimization.

- **Pure Strategy Evaluation.** At a minimum, calculate how well players will score if they play the game with simplistic strategies. Check *pure strategies*, in which a player always takes the same action regardless of the situation. If a player always evacuates regardless of the forecast, what is their average score? What if they never evacuate? If the goal is to teach players to balance risk with opportunity, neither pure strategy should score well. If one extreme really is the best behavior, then a game is probably not the best tool for that lesson. Games let players explore tradeoffs in different contexts; games are weak when the ideal strategy does not depend on context.
- **Machiavellian Playtesting.** Test the game with players who do not know the lessons you are trying to teach and do not care about learning the right lessons. Instruct them to just get the best scores possible. If they end up following best practice, then it's likely that the game is teaching the right lessons. If they end up deviating from best practice, the game is probably implicitly teaching the wrong lessons. Ask the playtesters what advice they would give to a new player who just wants to get a passing score—note whether their advice matches the learning objectives and focuses on the intended dilemmas. Do not actually give that advice to *students*! Assess whether a score-focused player will still learn the right lessons as a means of critically assessing the implicit lessons the game is teaching.
- **Computer-Aided Optimization.** Rapid play games are often amenable to being solved by a computer, but the overhead cost of building the infrastructure may make such methods less infeasible within a project's scope. For very simple games, a simple script may be able to exhaust all possible actions sequences, or a simple game theory model might be able to solve for optimal strategies (A.K. Dixit, 2011) (Nguyen, 2016). For moderate complexity games, machine intelligence, such as Monte Carlo tree search, can be added to iteratively optimize behaviors (D. Silver, 2018). For more complex games, modern machine learning methods, such as reinforcement learning, will often be able to solve the game (V. Mnih, 2013). If the best solutions found by the computer align with the behaviors you would like to see players embody, then the game is implicitly teaching the right lessons. If the best solutions deviate from best practice or are otherwise absurd, the game risks implicitly teaching the wrong lessons. Assess several locally optimal solutions found by the solver, not just the single globally optimal solution

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

– player may get stuck in local optima, and those local optima should, ideally, still teach a valid lesson.

In the radiological variant of Quake Responder, the player might be managing search and rescue (SAR) teams responding to city blocks with varying levels of need and radiological risk. Each turn represents one day of the response effort. Each turn, the player receives evidence about each site's need for SAR and radiological threat level. The player chooses which personnel will respond to each site and how long they will spend there. If a responder hits an unsafe threshold of radiation exposure, they are incapacitated, removing them from the pool of available responders for the rest of the game. Each time the player plays, the game varies the radiological threats at each site (sometimes generating quiet scenarios where there is no radiation at any site)—either randomly or according to a script that ensures a variety of experiences. Successful players must judge the evidence of radiation risk, rotate staff to keep each below the exposure threshold, and focus responders on the areas of greatest need. They might occasionally put responders at risk when lifesaving opportunities are high, and they will sometimes not respond to an area of need because the risks are too great.

Suppose the only scoring criterion in the radiological variant of Quake Responder were the number of lives saved. Such a scoring criteria is reasonable at face value, and it is how many experts describe the success of a response effort. However, score-focused playtesters and algorithms will discover that scores are maximized by ignoring radiological threat on the final turn. The only in-game penalty for exposing responders to harm is their absence on future turns, so there is no drawback to radiation exposure on the final turn. That loophole incentives behaviors contrary to the learning objectives; it might teach players incorrect lessons, or it might harm the credibility of other (valid) lessons the game conveys.

An easy fix would be to have a weighted scoring rule. A player is scored not only how lives saved but also on responders lost. Suppose the player loses 1 point per lost civilian and 1000 points per lost responder. That would solve the prior loophole, but likely creates a worse problem—a simple calculation will likely reveal that a player should never send a responder into a dangerous situation (at least, according to the scoring feedback they are getting). That incentive doesn't create a dilemma and doesn't force players to build a mental model about how much risk a given action will entail (and it may even teach an incorrect lesson). To serve the learning objective and build a flexible mental model of how to balance risk and reward, it is critical that some scenarios or some city sectors warrant putting responders at risk and some do not. If, however, we adjust the scoring weights so that pure strategies (always respond / never respond) are worse on average than a balanced approach, then the game forces players to grapple with the dilemma and build a mental model of the tradeoffs involved and learn what contextual details justify different responses.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

Pitfalls

- **Quiz**. A quiz has the ultimate degenerate strategy—just get every question right. A game that capitalizes on dilemmas and builds mental models should require considering context. A game with a dominant pure strategy is just a quiz.
- **Bloat**. Make sure that the tasks players are performing differ depending on scenario or situation. If the best action is always the same, eliminate that decision point to streamline the game or create additional scenarios to require player flexibility.
- **Irrelevant Mechanics**. Entertainment games require a great deal of work detecting and removing degenerate strategies through the process of 'balancing'. For serious games, the learning objective is the ultimate measure of success, not fairness or consistency. When adapting an entertainment game to a serious purpose, do not worry about undermining fairness and balance, as long as you still check for misleading degenerate strategies.
- **Silver Bullet**. Analyzing game strategies can consume arbitrarily much time and must be scoped realistically. If a player is only going to play a game a few times, all that matters is that pure strategies are punished. If they are going to play it 20 times, you need to put more work into ensuring that more complex emergent strategies are consistent with the learning objectives.

Technique 7: Report Relative Scores

A player may not be able to save all survivors, survive all disruptions, or achieve a 100% score on every scenario. Reliably perfect scores would be an unrealistic reflection of the real world and set false expectations. Furthermore, absolute scores are hard to interpret as being good or bad when scenario difficulty varies.

In Quake Responder, perhaps there are always 100 survivors in danger at the start of any scenario, but the difficulty of rescuing them varies based on procedurally generated factors. In one scenario, a player manages to rescue 45 survivors. Is that a good score or a bad score? Should they attempt to mimic that behavior in future games or try something different? The absolute score does not help the student build a mental model, undermining one of the primary benefits of an interactive instructional tool. However, if the game can compute, estimate, or lookup a pre-computed value indicating that the best possible raw score for that scenario is 50, the student can be given a relative score of 90% (45 out of 50). On the next scenario, the best possible raw score might be 90, making a raw outcome of 45 only a relative score of 50% (45 out of 90). The student's improvement trend can thus be tracked, even when individual scenarios difficulty varies greatly.

For some games, a solver might be able to compute the best possible score for an arbitrary scenario. In such cases, random scenarios can be generated with the solver running at game time. For more complex

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

games, it might be impossible or impractical to compute the best score for any given scenario. Instead, scenarios should be generated in advance (not at game time) so performance targets can be computed (automatically) or estimated (manually).

Pitfalls

- **Quiz**. The need for relative scores is a good sign that the game is not acting as a quiz and is helping the player build a flexible mental model.
- **Bloat**. Automatically computing the best possible score can be quite difficult for complex games. The need to be able to solve the game is a forcing function on keeping the game streamlined.
- **Silver Bullet**. Some tasks are always possible. If you are training players how to follow an incident response protocol, the best answer is to follow it 100% of the time. However, if a perfect score is always possible, reassess whether a game is the right tool or if the set of scenarios should be expanded to include more challenging situations. Often dilemmas emerge in the real world only when resources are overwhelmed, and giving a player a no-win scenario can draw out dilemmas in an otherwise simplistic task.

Technique 8: Build Player Decision Models

Some instructional settings have a known curriculum, and games and other instructional materials can target that curriculum when selecting scenarios, dilemmas, and other lessons to focus on. In other cases, the goal is to fill gaps in knowledge, in which case an important part of the instructional tool is to identify those gaps. Data from the employment of any instructional tool can assess which scenarios and topics gave students the most trouble. Data from a rapid play digital games can also be used to extrapolate typical behavior and predict performance on novel scenarios.

Techniques such as data clustering and Bayesian models can replicate typical behaviors of observed player data in novel situations (D. Ramachandran, 2007) (R.S. Sutton, 2018). The inferred model of player decision making allows an instructor to derive nightmare situations—scenarios with minimal difficulty and maximal chance of causing a typical student to fail. Those can augment a curriculum and focus students on current weaknesses. Decision models can also be used to generate a small set of scenarios that cover the space of possible scenarios, providing some assurance that the curriculum does not have major gaps in the range of scenarios it exposes students to.

Other machine-learning techniques such as inverse reinforcement learning can infer properties of player behavior (not just mimic it), such as identifying their priorities by inferring a reward function (B.D. Ziebart, 2008) (J. Choi, 2012) (M.C. Gombolay, 2016) (Jensen, 2019). Priorities from experts can be exposed and incorporated into the curriculum, while priorities of novices can be exposed to identify where they deviate from best practice.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

Pitfalls

- **Irrelevant Mechanics**. Entertainment games sometimes have a notion of ‘adaptive difficulty’, in which the game lowers its difficulty automatically in response to player failure. Building decision models is about creating custom scenarios to target player weaknesses.
- **Silver Bullet**. Training an accurate decision model requires sufficient data. It can be useful when assessing an entire class if the class is coming from a similar background and thus likely struggling with similar issues. It can suffer with diverse classes or when analyzing small data-sets from individual users. Building and tuning decision models requires development work, and it should not be lightly tacked onto a serious gaming project. With small classes, it can be more effective to just give instructors access to student game logs and let them draw their own conclusions about trends and shortfalls. Decisions models are most effective (and most easily justified) when scaling up to larger student populations.

Templates

The remaining sections examine five templates that are suitable for rapid play serious games. Unlike the earlier techniques, which are broadly applicable and largely cross compatible, these templates are more specialized to particular dilemmas, learning objectives, or the presence of existing materials. Each of these templates has been applied to multiple games serving multiple audiences and domains (Seater, 2019) (B. Soulliard, 2016) (H.J. Davison Reynolds, 2016) (R.M. Seater, 2018) (R. Seater, 2016) (M. Daggett, 2016) (Jensen, 2019).

Template 1: Time-Pressured Decision Quad

In many domains, key dilemmas arise from the tradeoff between information completeness and response timeliness. That dilemma can be mapped directly into a core game mechanic. The key in-game decision the player faces is whether to wait for more information (increasing response accuracy) or take decisive action (increase response effectiveness). It works best when the action being taken is irrevocable, expensive, or otherwise has a high consequence if taken unnecessarily or with the wrong target. It is effective at teaching players to judge the relative value of different types and combinations of information and building a mental model of how much information is needed to make an informed, but not overdue, decision.

For example, this template would be suitable for a dilemma about whether to evacuate in the face of a hurricane, whether national stockpiles should be deployed to preempt an infectious outbreak, or whether responders should pull out of an area with a hazard to the response team. It is less suited to sequences of small dilemmas like how to deal with press releases to the public, how to allocate scarce resources differently as an event unfolds, or how to recover in the aftermath of a disaster.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

The game interface typically has four main quadrants—two providing information and two allowing for player actions.

A) **Current Status.** Aggregate all the information the player has received so far, including the initial scenario priming and any information accumulated from previous turns. The information should be clearly presented to keep the game focused on the decision of *when* to act on incomplete available information, not testing memory or detective skills to find that information. If certain information arrives at known points in time, that schedule should be clearly summarized so the player is focused assessing the value of information they might wait for, not gambling on what information they will get if they wait.
B) **Turn History.** Summarize the actions taken and information acquired on each prior turn as a time-ordered list. Also include any turn limit information, so that a player can easily see how far they are into the scenario. The game should automatically end after a certain number of turns, when the response would no longer be effective, or to allow the player to express the belief that no action is required (e.g., in a quiet scenario).
C) **Primary Decision.** The player can commit to the primary decision, thereby ending the game early. The game ends since it is focused on knowing when there is enough information and what action that information requires, not on simulating the consequences of the action. This panel is also where the player indicates that they are not taking action and instead proceeding to the next turn.
D) **Supporting Decisions.** Allow the player to take minor actions, such as preparations (e.g. notifying the public, arranging contracts, staging supplies) or information-request actions (affecting what information the player will receive if they wait another turn to act). These decisions are, at most, a minor part of the final score; the primary score is a function of timeliness and accuracy of the primary decision.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

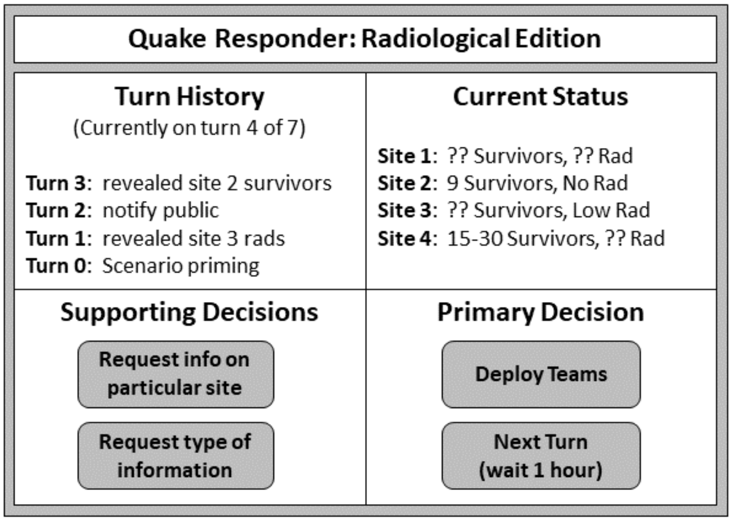

Figure 1. Time-Pressured Decision Template applied to the Quake Responder sample game. The player can see the actions taken and information revealed in prior turns (top left), the current information and what information is currently unknown (top right), minor responses available (bottom left), and the big decision of whether to act or wait for more information (bottom right).

Using this template, Quake Responder is focused on the decision of whether or not a response is appropriate and, if so, where it should focus. The key dilemma is whether to respond at all (are there survivors and is it safe?) and, if so, where to allocate teams (to maximize lifesaving and minimize exposure). Details about how the response will be executed are omitted as they are not critical to the core dilemma or core learning objective. As a supporting decision, the player can choose what information they will receive next turn—either specifying a site but not a type, or a type of information but not the site. Although the information request does not map directly to the real world, it serves the instructional goal of forcing the player to assess what type of information is most critical, along with the key lesson of judging how much information they need before committing to a plan of action.

Template 2: Perturbed Scientific Model

If established models already exist, they can often be leveraged into multi-player cooperative games focused on improving coordination (B. Soulliard, 2016). The existing model serves as most of the game mechanics, paired with rules for player actions to perturb exposed parameters and imperfectly inspect model elements. The basic turn structure is for all players to simultaneously submit actions, apply those actions to adjust model parameters, run the model for a fixed duration, then report back information from the model based on each player role. It is often useful to require that players pay for their actions from a common resource pool shared by all players. Communication between players is forbidden outside of limited in-game channels and may require expending shared resources. The key dilemma is

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

whether to act in the manner that is most efficient given current information or to spend time and resources coordinating with other actors to create a delayed but coordinated response.

For Quake Responder, suppose that a system dynamics model already existed for how populations in an earthquake zone move, share resources, degrade in health, and recover based on the presence of resources. The instructional goal is to improve coordination during the first 96 hours when communications and organizational infrastructure are not yet in place. In the game, the players represent search and rescue teams, medical teams, evacuation coordinators, and temporary housing providers. If well-coordinated, the players will create a pipeline moving survivors from one site to another as they are extracted, stabilized, transported, and housed. If uncoordinated, survivors might be recovered but fail to get medical attention at their location, or survivors might be evacuated to a location with inadequate housing.

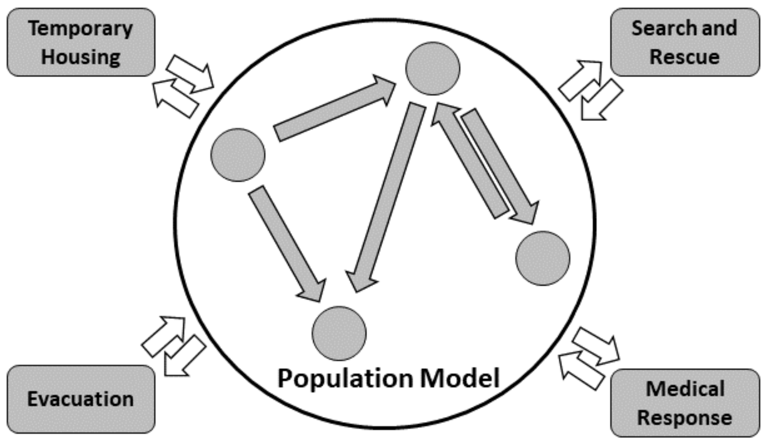

Figure 2. Perturbed Scientific Model Template applied to the Quake Responder sample game, with four disjoint team coordinating to serve a common population. The model can simulation baseline outcomes, or it can be perturbed by player actions mid-run to simulate the effects of intervention.

Players of Quake Responder can perturb the underlying model by injecting resources or raising the health parameter of survivors at a particular location. The model then simulates the effects of those actions and emergently demonstrates the efficacy of coordination. Each player can observe a limited portion of the model. For example, the search and rescue team can use helicopters to observe the number of survivors at each location, which is information that the evacuation coordination players needs to allocate their personnel. The medical response team has information about the survival rates of survivors extracted from different sites, which is information that the search and rescue team needs to prioritize sites. Players can only communicate by sending a text message through the game interface,

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

but doing so reduces their ability to take any other action that turn. The primary dilemma faced every turn is whether to communicate, how much to communicate, with whom to communicate, and how to allocate scarce resources as a result of that communication. In each scenario, the game randomly varies the disaster scale, the number of coordinating teams, and the cost of communication, thereby forcing players to balance coordination and efficiency differently in different contexts.

Template 3: Replicated Tool Interface

When software tools already exist for the task being trained, or for closely related tasks, their interfaces can be leveraged to guide the design of the game and improve its effectiveness (H.J. Davison Reynolds, 2016). If possible, use the actual tools and instrument them with data from the procedurally generated scenarios. The game structure will have three parts—(1) scenario priming, (2) interaction with an imitation of existing tools, and (3) a decision input and score summary screen. The bulk of the game is spent in the second part (interaction with actual tools), but the player will have to make a higher-level strategic decision (the dilemma) based on those interactions. The player practices the individual tools but is trained and evaluated on applying those raw skills to making a higher-level decision in a particular scenario context. For example, the player's tasking would not be to determine wave height forecasts but rather to determine which districts should prepare for evacuation and when the final call should be made.

The technical cost of interfacing with an existing tool can be high. When it is feasible, there are a number of design benefits beyond tool-specific training. The information available through the existing interface will help manage bloat and scope creep, and the data interfaces will help to define the key elements of procedural scenario generation. The design of the original interface will often do a lot of the work of identifying information relevant to a key decision and focusing the problem on a relevant subset of the world. Matching or reusing a real tool interface can help translate the lessons to the field by training the player in an environment that matches real operating conditions.

For Quake Responder, suppose a database tool already existed for building response teams and schedules. A player first sees background information on a particular event, then uses the real tool to build a response plan, then submits an export of that plan to an adjudicator that scores it against the scenario priorities. If, instead, the existing tool were a radiological risk assessment calculator, the player might be given access to the real tool during gameplay, but still enter their final decisions for how to allocate teams between sites via the game interface. In both versions, the players practice use of the tool in the context of the more judgement-based dilemmas the game presents.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

Template 4: Murphy's Law Red Teaming

A big benefit of professional military wargames is the inclusion of an active adversary—the red team—who can respond intelligently and creatively to the blue team (Perla, 2012) (P. Harrigan, 2016). The participants are being trained to be better at the blue team's task, and the opposing red team serves two main instructional roles—(1) it gives the blue team a more realistic opposition who is creative yet makes mistakes, and (2) it gives blue team members a chance to reverse roles and see the problem from another perspective. Thinking like the enemy and dealing with a thinking enemy are both valuable to learning.

Emergency management is not usually a situation with an active or intelligent adversary. Even if the event was an act of terrorism, the perpetrators are generally long gone by the time the response begins, and the response is not being actively thwarted by their efforts. A non-adversarial game uses procedurally generated scenarios to surprise the player and scientific models to create a realistic feedback. Often, those are enough to meet the learning objectives. However, if the learning objectives emphasize improvisation, flexibility, and adaptation then the in-game situations need to be more dynamic, unpredictable, and responsive. A human adversary often provides those properties far better than an automated model. The design challenge is to incorporate an active adversary into a non-adversarial domain in order to gain those benefits without losing realism or credibility.

When modeling a domain without an active adversary, an adversarial role can be created by treating Nature or Murphy's Law as the adversary. A single-player or cooperative game can be transformed into an adversarial competitive game by putting a red team in charge of scenario creation and evolution. Random elements of the game, both in scenario generation or during gameplay, are instead controlled by the red team. Give the red team an abstract "Murphy's Law" budget to spend to affect otherwise random elements. The cost to alter a random result is inversely proportional to the rarity or certainty of that event or phenomenon. The red team can gain points by making events go in favor of the blue team, presumably when it will least benefit them.

In Quake Responder, add a red team in the form of Nature. Nature's job it so make the disaster the "perfect storm" of bad luck by spending an 'unlikelihood' budget to adjust scenario parameters. For a low cost, they set the time of day of the event to be rush hour. For a high cost, they have the event happen during a holiday with a larger number of tourists present. After the blue team proposes an evacuation plan, Nature pays a low cost to add severe thunderstorms. After the blue team deploys temporary housing, Nature pays a high cost to destabilize the nearby nuclear plant. After the blue team deploys teams to stabilize the reactor, Nature recovers some of their budget by declaring the instability to be a false positive posing no health threat, wasting blue team's time and resources. Nature spends its

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

final budget to partially distort information about which areas lack clean water, hoping to induce mistrust in all of the data the blue team receives.

When students are pitted against an active red team, there are instructional benefits and drawbacks. They will encounter more difficult and surprising situations, forcing them to adapt and not fall into overly formulaic patterns. However, the instructional experience will lose consistency, comparability, and coverage of the material. Two students might face the same adversary and encounter very different levels of difficulty depending on how well the red team played that particular game. They might face different challenges and learn different lessons than other players. Those lessons might be individually richer but less measurable. An active adversary provides a richer experience but a poorer assessment tool.

Template 5: Layer of Fog, Friction, and Chance

Clausewitz (Clausewitz, 1832) described three properties that differentiate ideal war from real war—fog (lack of perfect information), friction (lack of perfect control), and chance (lack of accurate forecasting). *Fog* is when your information is incomplete or inaccurate, *friction* is when your actions are not precisely carried out, and *chance* is when the future cannot be fully predicted from the present or from a proposed action. The same principles are critical to real-world emergency management and are often omitted from instructional material (P. Harrigan, 2016).

Given an existing game or model that lacks those aspects, a new game can be derived by adding an interference layer between the game and the player. The interference layer randomly hides, delays, or corrupts information coming from the underlying game to the player (fog) and/or randomly drops, delays, or corrupts instructions sent from the player to the game (friction). Even low levels of fog and friction can dramatically change the optimal strategies in a game, making them better reflect the real world. The presence of fog and friction will place greater emphasis on robustness and flexibility in the face of chance and less emphasis on optimization and precision.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

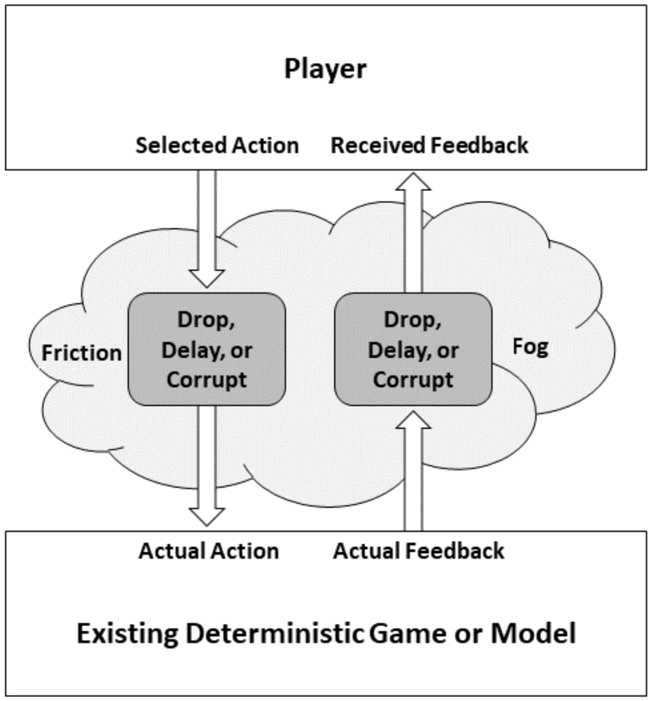

Figure 3. Schema for inserting a layer of fog, friction, and chance between the player and an existing game. Fog is a limitation of what is known about the state of the world (e.g., masking or randomly distorting the truth). Friction is a limitation on the ability to affect the state of the world (e.g. delaying or distorting commands given).

In Quake Responder, a player's command to allocate resources to a particular site could be randomly delayed by 0–3 turns. Now, instead of creating a carefully orchestrated sequence of actions, the player needs to create a more robust plan that will not fall apart if it becomes partially desynchronized. The player will also need a greater emphasis on adapting their plan, even when the earlier plan was a good one, thereby more directly measuring their flexibility and adaptability.

As with transparency of information (Technique 5), fog and friction can be treated as a difficulty knob—increase it only after the student is familiar with the unaltered game. Basic principles are best taught with a clear and non-random feedback, whereas advanced skills and methods require that the uncertainties of fog and friction be faced. As with all methods, keep an eye on the learning objectives and the key dilemmas being modeled to judge how much (if any) fog, friction, and chance are appropriate.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

Takeaways

Rapid play serious games are a powerful instructional tool that, like all tools, can fail to be of value or become prohibitively expensive if pursued naïvely.

1) **Quiz Pitfall.** Is your game just testing factual knowledge that is independent of context and does not require a flexible mental model?
2) **Bloat Pitfall.** Have you included details in the name of realism that do not serve the learning objectives? Have you distracted students from the learning objectives by over-emphasizing secondary aspects of the domain?
3) **Irrelevant Mechanics Pitfall.** Have you included mechanics because they are in similar entertainment games without assessing their relevance to the learning objectives and impact on player incentives?
4) **Silver Bullet.** Are you building a game when a different instructional medium would be more appropriate? Are you pairing rapid play games with complementary instructional methods?

The risks of such failures can be greatly reduced by applying the eight techniques described above. Those techniques complement and concretize principles of game design (D. DellaVolpe, 2013) (Burgun, Clockwork Game Design, 2015) (Schell, 2014) and of education (Kang, 2016) (T.C. Toppino, 2014) into the context of instructional games.

1) **Focus on Dilemmas** to avoid including extraneous or distracting details.
2) **Include Multiple Scoring Criteria** to focus the player on the actions that trade off the aspects of those dilemmas.
3) **Procedurally Generate Scenarios** so players must respond differently to the same dilemma in different contexts.
4) **Include Quiet Scenarios** to create uncertainty and mitigate the artificiality of the instructional setting.
5) **Stretch the Reality of Rarity, Transparency, Authority, and Restrictions** to force players to build richer mental models and understand the purpose of doctrine and policy.
6) **Analyze Degenerate Strategies**, especially pure strategies, using computer-aided analysis when possible to ensure that the game is teaching the right implicit lessons.
7) **Report Relative Scores** to make easy and hard scenarios comparable and to convey a realistic standard of success.
8) **Build Player Decision Models** to identify skill gaps and fill curriculum gaps.

Consider using one of these templates as a starting point:

1) **Time-Pressured Decision Quads** are useful when the key dilemma is timeliness vs. accuracy.
2) **Perturbed Scientific Models** are useful when the key dilemma involves coordinating activity with limited communication.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

3) **Replicate Tool Interfaces** to teach strategic lessons in the context of familiar toolsets.
4) **Murphy's Law Red Teaming** can add an active adversary to a non-adversarial situation.
5) **A Layer of Fog, Friction, and Chance** can force a player to think more about flexibility and robustness vs. optimization and precision.

When employed carefully, rapid play games are a useful tool to supplement other instructional methods. A rapid play game's strengths stem from allowing a student to play many different scenarios in rapid succession, quickly seeing the consequences of their actions, and building mental models about how those consequences vary with context. These strengths are especially relevant when the instructional goals include flexibility, judgement under pressure, and preparedness for events where policy and procedure do not completely dictate response.

References

A.K. Dixit, B. N. (2011). *The Art of Strategy: A Game Theorist's Guide to Success in Business and Life.* New York: W.W. Norton & Co.

Adams, E. (2013). *Fundamentals of Game Design, Third Edition.* Thousand Oaks, CA: New Riders.

B. Soulliard, H. D. (2016). A Simulation Capability for Training Efficient Air Traffic Management. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference.* Orlando, FL.

B. Vogt, M. M. (2015). Early Synthetic Prototyping: When We Build It, Will They Come? *Interservice/Industry Training, Simulation, and Education Conference.* Orlando, FL.

B.D. Ziebart, A. M. (2008). Maximum Entropy Inverse Reinforcement Learning. *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence.*

Burgun, K. (2012). *Game Design Theory.* A K PETERS.

Burgun, K. (2015). *Clockwork Game Design.* Focal Press.

Chekhov's Gun. (2020). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Chekhov%27s_gun

Clausewitz, C. V. (1832). *On War.*

Conde-Pumpido, T. d. (2017). Persuasive Gaming: Identifying the different types of persuasion through games. *International Journal of Serious Games, 4*(1).

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience.*

D. DellaVolpe, R. B. (2013). *War Gamers' Handbook: A Guide for Professional War Gamers.* Newport, Rhode Island: U.S. Naval War College.

D. Ramachandran, E. A. (2007). Bayesian Inverse Reinforcement Learning. *International Joint Conferences on Artificial Intelligence (IJCAI).*

D. Silver, T. H. (2018). A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go through Self-Play. *Science., 362*(6419), 1140-1144.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

E. A. Kensinger, S. C. (2003). Effect of Negative Emotional Content on Working Memory and Long-Term Memory. *Emotion*, 378–393.

E. Grois, W. H. (1998). Bayesian Network Models for Generation of Crisis Management Training Scenarios. *The Tenth Conference on Innovative Applications of Artificial Intelligence (IAAI-98).* Madison, WI.

H.J. Davison Reynolds, M. P. (2016). Improving the Emergency Manager's Hurricane Evacuation Decision Making through Serious Gaming. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference.* Orlando, FL.

Hattie, J. (2012). *Visible Learning for Teachers: Maximizing Impact on Learning.* Routledge.

J. Choi, K. K. (2012). Nonparametric Bayesian Inverse Reinforcement Learning for Multiple Reward Functions. *Advances in Neural Information Processing Systems 25 (NIPS).*

Jensen, R. (2019). *Strike Group Defender*. Retrieved from MIT Lincoln Laboratory: https://www.ll.mit.edu/r-d/projects/strike-group-defender

K. L. Kettle, G. H. (2010). *Motivation by Anticipation: Expecting Rapid Deedback Enhances Performance.* Association for Psychological Science.

Kang, S. (2016). Spaced Repetition Promotes Efficient and Effective Learning: Policy Implications for Instruction. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 12-19.

Klein, G. A. (1998). *Sources of Power : How People Make Decisions.* Cambridge, MA: MIT Press.

M. Daggett, K. O. (2016). Predicting Team Performance Through Human Behavioral Sensing and Quantitative Workflow Instrumentation. *Proceedings of the AHFE 2016 International Conference on Human Factors and System Interaction: Advances in Human Factors and System Interactions.* Springer.

M.C. Gombolay, R. J. (2016). Apprenticeship Scheduling: Learning to Schedule from Human Experts. *Proceedings of the International Joint Conference on Artificial Intelligence.*

Nguyen, T. (2016). Serious Games. *Significance, 14*(6).

P. Harrigan, M. K. (2016). *Zones of Control: Perspectives on Wargaming.* Cambridge, Massachusetts: MIT Press.

P.M. Kato, S. C. (2008). A Video Game Improves Behavioral Outcomes in Adolescents and Young Adults With Cancer: A Randomized Trial. *Pediatrics, 122*(2), e305-e317.

Page, E. H. (2019). *Modeling and Simulation, Experimentation, and Wargaming: Assessing a Common Landscape.* The Journal of Defense Modeling & Simulation.

Perla, P. (2012). *Peter Perla's The Art of Wargaming: A Guide for Professionals and Hobbyists.* Annapolis, Md: U.S. Naval Institute Press.

R. Seater, C. R. (2016). *Skill Transfer and Virtual Training for IND Response Decision Making: Game Design for Disaster Response Training.* Lexington, MA: MIT Lincoln Laboratory.

R.M. Seater, J. K. (2018). Rapid-Play Games for Evaluating Future Technology. *Proceedings of the 7th International Conference of the Games and Learning Alliance (GALA).* Palermo, Italy.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.

R.S. Sutton, A. B. (2018). *Reinforcement Learning: An Introduction.* Cambridge, MA: MIT Press.

S. Garozzo, S. S. (2015). *Emphasizing Player Skill and Strategic Choices in CS:GO Level Design.* San Francisco, CA: GDC: Game Developers Conference. Retrieved from https://www.gdcvault.com/play/1021868/Community-Level-Design-for-Competitive

S.A. Yourstone, H. K. (2008). Classroom Questioning with Immediate Electronic Response: Do Clickers Improve Learning? *Decision Science: Journal of Innovative Education*.

Schell, J. (2014). *The Art of Game Design: A Book of Lenses, Second Edition.* CRC Press.

Seater, R. (2019). Using Rapid-Play Serious Games and Crowdsourcing to Inform Technology Evaluation and Research Prioritization. *Massachusetts Institute of Technology Lincoln Laboratory Journal, 23*(1).

T.C. Toppino, E. G. (2014). About Practice: Repetition, Spacing, and Abstraction. *The Psychology of Learning and Motivation*, 113-189.

T.M. Connolly, E. B. (2012). A Systematic Literature Review of Empirical Evidence on Computer Games and Serious Games. *Computers & Education, 59*(2), 661–686.

The Law of Conservation of Detail. (n.d.). Retrieved from TV Tropes: https://tvtropes.org/pmwiki/pmwiki.php/Main/TheLawOfConservationOfDetail

V. Mnih, K. K. (2013). Playing Atari with Deep Reinforcement Learning. *Computing Research Repository (CoRR)*.

Zhonggen, Y. (2019). A Meta-Analysis of Use of Serious Games in Education over a Decade. *International Journal of Computer Games Technology*.

This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering.