Proceedings, The Eighth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment

If Not Now, Where? Time and Space Equivalency in Strategy Games

Baylor Wetzel[1] and **Kyle Anderson**[2] and **Maria Gini**[1] and **Wilma Koutstaal**[3]
[1]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455
{wetzel,gini@cs.umn.edu}
[2] Shikigami Games, {kyle@ShikigamiGames.com}
[3]Department of Psychology, University of Minnesota, Minneapolis, MN 55455
{kouts003@umn.edu}

Abstract

Spatiotemporal reasoning is a fundamental contributor to effective problem solving. In an effort to design better problem-solving agents, we examined and evaluated the strategies that humans use to solve Tower Defense puzzles, a complex and popular class of real-time strategy games. A consistent and unexpected finding was that humans frequently treated time and space as equivalent. Players stated temporal goals but solved spatial problems. An analysis of human data and computer simulations showed that re-representing temporal problems as spatial problems was beneficial, but treating the two separately can lead to higher scores. The work presented here holds several possibilities for level designers and others who design and analyze maps and spatial arrangements for domains requiring strategic reasoning.

Introduction

While computers have exceeded humans at many types of reasoning tasks, they still lag behind humans in complex temporal and spatial reasoning tasks such as the game of Go and real time strategy (RTS) games (Buro 2003).

An interesting and previously unstudied domain of this type is Tower Defense (TD) games. In TD the player attempts to stop enemies from moving across a map by placing guard towers at strategic locations. TD combines the complexity of spatiotemporal reasoning found in RTS games with a simple environment that is deterministic, non-adversarial and perfect information. TD provides an arena in which to study spatiotemporal reasoning, problem and process representation, and transferring expertise across maps.

We created a research version, called *Gopher TD*, of a popular TD game and instrumented it to study how humans represented, reasoned about, and solved the spatiotemporal reasoning problems presented by TD. Although TD initially appears to be a predominantly spatial reasoning problem, solution quality depends heavily on subtle and nonobvious temporal considerations.

One notable observation, derived from observing humans attempting multiple TD problems under a think-aloud protocol, is that subjects often treated time and space as interchangeable. This is consistent with findings and theoretical accounts of other researchers with respect to language (Boroditsky 2000) and motor actions (Miles et al. 2010) and might be innate (Mandler 2012). We are unaware of any research showing how this conflation affects task performance. We found that, compared to a random baseline, re-representing time as space performed quite well. However, by explicitly decoupling the two, experienced players were often able to earn higher scores.

In this paper we discuss the concept of space-time conflation, introduce several spatiotemporal strategies used by human subjects and the effectiveness, applicability and limitations of those strategies we implemented in various AIs. We first introduce the domain and then discuss the early study which led to a catalogue of strategies and representations that informed the construction of the AIs. The main part of the paper presents the results of a recent study with human subjects in which we evaluated a subset of the strategies used in our initial study and compared their performance to the performance of the AIs.

The contributions of this paper include introducing the TD domain as an environment for studying spatiotemporal reasoning, showing the feasibility of using spatial reasoning to solve temporal problems, and illustrating how decoupling of time and space can, counterintuitively, lead to a more advanced manipulation of time.

Related Work

People, when faced with spatiotemporal decision problems, such as planning a shopping trip, construct mental representations that simplify reality and that are specific to the task at hand (Arentze, Dellaert, and Timmermans 2008).

Mental model theory (Johnson-Laird 1983) states that individuals mentally simulate the effect of actions before executing them. Previous work has studied spatial representations used by people, such as the mental spatial maps constructed for navigation tasks (Kuipers 1978), and representations used in naive physics models for predicting the outcome of an action (Friedman and Forbus 2009).

Research suggests that simpler mental models are preferred over more complex models for making deductions because the inference process is easier (Johnson-Laird 2001; Jahn and Knauff 1998). Problems that require multiple mental models are harder for people. Specifically, when people use temporal relations to assess the relative order of events, having a single mental model reduces processing time and increases accuracy (Schaeken, Johnson-Laird, and d'Ydewalle 1996). Similar effects have been shown for problems that require both temporal and spatial reasoning (Vandierendonck and Vooght 1997), creating strong support for how humans use mental models.

When people think about time, they use whatever spatial representation is cognitively available to them. Time typically is mapped to one dimensional space, for instance, moving from left to right, or from back to front, to represent the passage of time (Miles et al. 2010; Casasanto and Boroditsky 2008). Time is conflated with one-dimensional space, but in this paper we address how humans reason about temporal effects of actions performed in a two dimensional space. Specifically, we examine how humans reason about the motions of objects in 2D that move at different speeds.

Understanding mental models is essential to create cognitive models, such as ACT-R (Anderson et al. 2004), which models low-level operations and mental resources and can be used to predict task difficulty and performance time. An ACT-R model of simple human spatial relational reasoning has been developed (Ragni and Brüssow 2011). In this study we are not aiming at creating a model of human spatiotemporal reasoning, but by understanding how humans handle spatiotemporal problems we provide the foundations for designing agents capable of replicating human performance.

Our work differs from computer-based complex scenarios or microworlds, used to study dynamic decision making and complex problem solving (Gonzalez, Vanyukov, and Martin 2005), in which people actively interact with an unknown (nontransparent) system of many highly interrelated variables so as to actively generate knowledge to achieve certain goals. TD problems contain no "hidden variables" and the potential interactions of space and time – though sometimes nonobvious and subtle – are open for observation.

In addition to providing insights on human reasoning, and providing strategies for implementation in an agent, computational models of space and time have many practical applications, such as urban planning.

Tower Defense

Tower defense (TD) games ask players to protect a base by organizing the base's defenses. The player is given a map (typically a simple maze) showing the path the enemies (generically referred to as creeps) will follow through the maze and a budget for purchasing defense towers to be placed on the walls of the maze.

The player cannot directly interact with the creeps. The only choices the player has are which types of towers to use and where to place them. Towers have a fixed rate of fire, do a fixed amount of damage per shot and never miss. Creeps move at a fixed speed and never veer from the path. The score in the game is based on the number of creeps stopped. Although the creeps are nominally the enemy, they are deterministic and do not react to the player. As such, TD is technically a puzzle, not an adversarial game.

A tower is only active when a creep is in its range. Because a tower's rate of fire and damage per shot are fixed, the only way to increase a tower's effectiveness is to increase the time the tower is active, which means increasing the time creeps are in a tower's range. Although placing a tower is a spatial decision, the ultimate goal is to maximize a temporal variable, the amount of time a tower fires. This points to the dependency between time and space, which is the subject of this study.

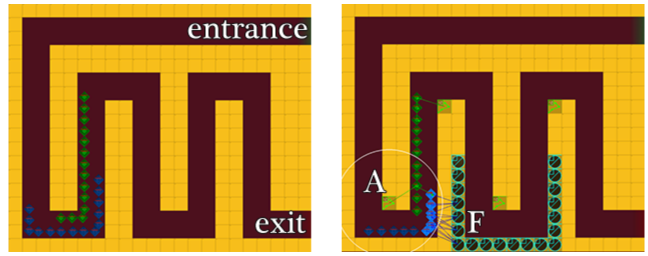

Figure 1: *Left*: Two lines of creeps move at a constant speed along a pre-determined path. *Right*: Attack towers (*A*) damage creeps while freezing towers (*F*) slow them down. The white circle shows the area in the tower's range area.

Experimental Setup

Studies were conducted with *GopherTD*, which we based on the commercial game Vector TD. GopherTD captured subject actions, timings, solutions and scores. Subject thought processes were recorded under a think aloud protocol (Lewis 1982) and participants were interviewed after completion of the testing phase. The data was compiled and qualitatively scored and categorized by multiple reviewers. These data were used to understand and classify the strategies used by the subjects and later to inspire the design of several AIs.

GopherTD contains 16 maps. Each path on the map has 14 creeps moving along it. There are different types of towers and those types differ in cost, attack power, speed, and range. Towers are placed on a grid (the gold squares in Figure 1) but creeps move continuously along a pre-set path (maroon squares in Figure 1). All towers attack the closest creep, focusing on that creep until it leaves range.

Subjects were grouped, based on self-report, into novice players (no knowledge of TD puzzles) and experienced players (familiar with TD rules and concepts). No subject had prior experience with GopherTD.

Study 1 (n=38) focused on collecting data on subject representations, strategies and problem solving processes. Unexpectedly, no two players used the same sets of strategies. Study 1 showed the breadth of the domain but the sparse nature of the data made in-depth analysis difficult.

Strategies depend in part on which towers the subject chooses. In study 2 (n=13; 4 novice, 9 experienced), all subjects were required to use the same set of towers, constraining the problem enough to make in-depth analysis possible. Only two types of towers were used. Attack towers fire 10 times a second and can attack 3 targets simultaneously. Freezing towers fire once every two seconds, attack 4 targets simultaneously and cause affected creeps to move at half speed for two seconds. Subjects had 4 attack towers and 20 freezing towers. Selected towers and quantities were

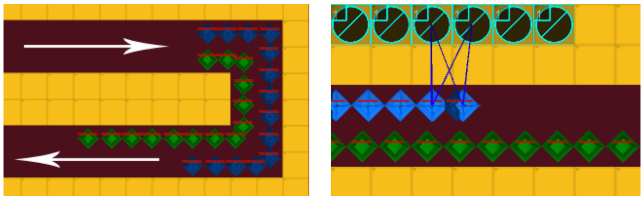

Figure 2: Objects on paths fall out of synchronization. *Left*: Outer line has longer path around a corner. *Right*: Freezing one line, ignoring the other.

based on those used by top subjects in study 1 and independently verified through simulation.

In the training phase, subjects first played each map three times. In the test phase, subjects played each map one more time. Test scores and solutions were recorded.

Study 1: Results

Below are the main strategies subjects used in Study 1.

Temporal Strategy: Slow In Range. All subjects placed freezing towers such that creeps were slowed while moving through the attack tower's range. We refer to this strategy as SLOWINRANGE. Unless otherwise mentioned, it should be assumed that all subjects used this strategy in combination with their other strategies.

Spatial Strategy: Advanced Maximum Usable Range. All subjects, regardless of strategy, tried to maximize the amount of path area inside a tower's range, called the usable range. The MAXIMUMUSABLERANGE (MUR) strategy focuses only on usable range. No human player used MUR by itself, but many, especially the less experienced players, used the ADVANCEDMAXIMUMUSABLERANGE (AMUR) strategy which combines MUR with SLOWINRANGE.

Temporal: Path Sync-Based. Although the two lines of creeps normally move through the map side by side, several subjects noted that they would sometimes fall out of sync. This can be caused by the map geometry. When going around a corner, the creeps on the outside path travel a longer distance and therefore fall behind those on the inner path (Figure 2, left). We refer to this gap as the path synchronization gap and placing towers where the natural gap is large as the EXPLOITGEOMETRY strategy. Subjects who used this strategy argued that, by having fewer creeps in range at one time, the tower was less likely to be overwhelmed, and by stretching out the arrival rate of the creeps, the tower could be active for longer.

We refer to the strategy of using freezing towers to create or enhance this gap as differential slowing, or DIFFSLOW (Figure 2, right). In some cases, it is possible to completely separate the two lines, doubling the activation time (and thus effectiveness) of the tower. Using DIFFSLOW and EXPLOITGEOMETRY together delays an already delayed line.

The path synchronization gap-based strategies are harder to execute than AMUR and require several optimization decisions. Attack towers must be placed after creeps have been separated, which can rule out positions with significantly more usable range. DIFFSLOW requires placing freezing towers outside the attack towers' range, leaving fewer resources for the SLOWINRANGE strategy (the ratio of freezing towers allocated to DIFFSLOW vs. SLOWINRANGE varied by subject). It is difficult to execute path synchronization gap-based strategies if attack towers are spread out. SLOWINRANGE often undoes the effect of the other strategies, making it hard to predict the creep line relationships for later towers. DIFFSLOW cannot be used if there is no position that can target only the desired line.

Subject Descriptions

A notable feature of subject descriptions is that language was predominantly spatial. People who used the AMUR strategy typically used terms such as, "I want to cover as many maroon squares as possible" (path area was maroon), "I chose corners because they overlook the most [path] area," and "I want to keep the creeps slow while they are in range." Those who used temporal strategies used a mixture of temporal and spatial language. They would say things such as, "I look for areas where the creeps are separated," "I'm going to pull this line so that they fall behind" and "I want to give the tower more time to focus on each creep."

It is important to note that problem solving reflected a combination of simple strategies rather than a single comprehensive strategy. Most subjects drew from the same set of strategies but combined them in ways that resulted in a unique play style.

We observed six (non-exclusive) classes of strategies: positional (placement choices of towers on the map and with respect to one another), tower selection (preference for particular tower types or combinations), target selection (placement of towers with respect to which creeps they would attack), tower distribution (how towers were placed with respect to one another, e.g., single kill zone, spread out), emotional-motivational (e.g., place towers near exit to minimize regret) and aesthetics (e.g., preference for symmetrical tower placements).

Strategies ranged in complexity from rules like the ones above to patterns (always place towers in a 4×2 pattern) to biases (prefer corners, avoid single-sided interior corners). Multiple strategies could be used concurrently. For example, DIFFSLOW was always paired with the AMUR strategy and instantiated as a compromise between maximizing path gap and usable range.

Strategies were often described in spatial terms. For example, in the target selection strategy in which towers are placed near each other so that they focus on the same target, subjects would explain their reasoning as "I am trying to kill the creeps as fast as possible [temporal goal] and because the towers select the creep that is closest to them, I try to put the towers as near each other as possible [spatial action] so that they select the same creep."

Study 2: Strategy Evaluation

Although every map had a substantially different layout, subjects tended to categorize maps based on how amenable they were to their preferred strategies. The path synchronization-based strategies try to decrease window

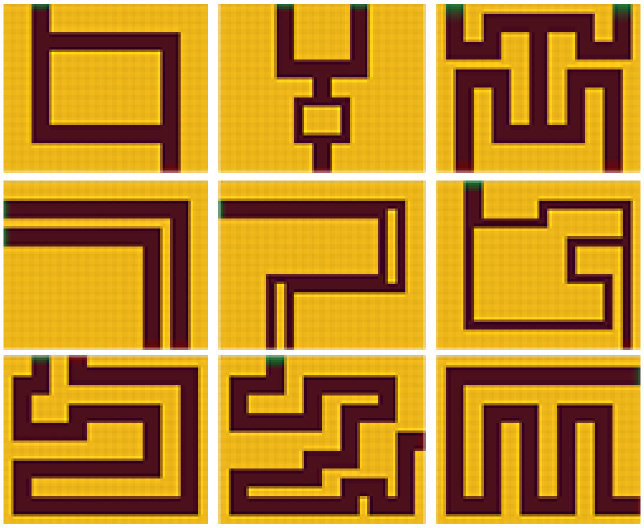

Figure 3: *Top*: NoGap maps. Paths have the same length. *Middle*: PersistentGap maps. One path is longer than the other. *Bottom*: DynamicGap maps. The outer path around corners is longer, causing creeps on it to fall behind, but later corners in the opposite direction allow creeps to catch up.

density by stretching the two creep lines using spatial geometry or freezing towers. The degree to which one line has fallen behind another is the path gap. Using type and frequency of path gaps, we grouped the maps into the categories of *NoGap* maps, *PersistentGap* maps and *DynamicGap* maps. NoGap maps cannot use the EXPLOIT GEOMETRY strategy. PersistentGap maps have one line of creeps that is consistently behind the other. DynamicGap maps have corners that cancel each other out, minimizing the number of places to apply EXPLOITGEOMETRY and undoing the effects of DIFFSLOW.

Study 2: Results

Preliminary data screening showed that three maps yielded very high (ceiling performance) scores, one map led everyone to use basically the same solution and one map was qualitatively highly similar to another map. We focus on the 9 standard maps (three in each of the three categories) that yielded a range of scores and strategies across participants. Average scores for the AI and Human player types for the three map types are given in Table 1. The AI was run 100 times and the mean score reported. Score is the number of creeps killed and ranges from 0 to 28.

Experience Level: Novice vs. experienced. We first examined skill differences between novice and experienced players to determine whether the domain is complex enough to require experience but solvable by experienced humans. A 3 (Map Type: NoGap, PersistentGap, DynamicGap) $\times$ 3 (Map Number: 1, 2 or 3) $\times$ 2 (Experience Level: Novice, Experienced) mixed factor analysis of variance (ANOVA), with map type and map number as within-subjects factors and experience level as a between-subjects factor, showed a significant main effect of experience level: $F(1, 11) = 7.48$, $p = .019$, $MSE = 21.51$, with experienced players ($M = 23.46$) outperforming novices ($M = 20.92$).

Map Type: NoGap, PersistentGap, DynamicGap. To determine whether key structural properties of the map influenced the task difficulty as reflected by scores, we examined performance as a function of map type. The above $3 \times 3 \times 2$ ANOVA revealed a significant main effect of map type: $F(2, 22) = 52.95$, $p < .001$, $MSE = 11.71$, with means 19.85, 19.54 and 27.17 for the three map types, respectively.

Experience Level $\times$ Map Type. To determine whether or not novices were uniformly worse than experienced players, we looked at the interaction of experience level and map type. The above $3 \times 3 \times 2$ ANOVA showed a significant map type $\times$ experience level interaction, $F(2, 22) = 3.59$, $p = .046$, $MSE = 11.71$. Experienced players outperformed novices on NoGap maps (20.70 vs. 19.00) and PersistentGap maps (22.07 vs. 17.00) but not on DynamicGap maps (27.59 vs. 26.75). The experience advantage was greatest for PersistentGap maps.

Excluding DynamicGap Maps. All subjects showed near perfect performance on the DynamicGap maps. To determine whether this masked other effects, we reanalyzed the data excluding these maps. A 2 (Map Type: NoGap, PersistentGap) $\times$ 3 (Map Number: 1, 2, 3) $\times$ 2 (Experience Level: Novice, Experienced) mixed factor ANOVA, with map type and map number as within-subjects factors and experience level as a between-subjects factor again showed a significant main effect of experience level: $F(1, 11) = 5.76$, $p = .035$, $MSE = 33.15$, with experienced players ($M = 21.39$) outperforming novices ($M = 18.00$). There was now no main effect of map type, $F < 1$, but again a strong trend toward a map type $\times$ experience level interaction, $F(1, 11) = 4.39$, $p = .06$, $MSE = 10.74$. Experienced players outperformed novices on NoGap maps (20.70 vs. 19.00) and PersistentGap maps (22.07 vs. 17.00) but the experience advantage was greatest for the PersistentGap maps (1.70 vs. 5.07).

Human Strategy. All novices (100%) and one third (33%) of experienced players relied solely on the AMUR strategy. This was the only strategy used by these players. Two-thirds (66%) of experienced players frequently used DIFFSLOW, either with (33%) or without (33%) EXPLOITGEOMETRY. All subjects used either AMUR (54%) or DIFFSLOW (46%).

Subjects were classified into two groups based on whether they used spatial strategies (AMUR, $n = 7$) or temporal strategies (DIFFSLOW, DIFFSLOW + EXPLOITGEOMETRY, $n = 6$). A 2 (Strategy: Spatial, Temporal) $\times$ 2 (Map Type: NoGap, PersistentGap) $\times$ 3 (Map Number: 1, 2, 3) mixed factor ANOVA, with strategy as a between-subjects factor and map type and map number as within-subjects factors showed a significant effect of strategy, $F(1, 11) = 11.73$, $p = .006$, $MSE = 24.44$, with spatial ($M = 18.57$) scoring, on average, significantly lower than temporal ($M = 22.42$). There was also a strong trend toward a strategy $\times$ map type interaction, $F(1, 11) = 3.91$, $p = .07$, $MSE = 11.09$, reflecting a smaller advantage for temporal strategies on the NoGap maps (spatial = 19.10, temporal = 21.44, advantage for temporal = 2.34) than on the PersistentGap maps (spatial = 18.05, temporal = 23.39, advantage for temporal = 5.34).

		NoGap			PersistentGap			DynamicGap		
Strategy	Player	N1	N2	N3	P1	P2	P3	D1	D2	D3
RANDOM	AI	3.86	8.18	16.60	4.74	9.82	14.25	26.84	23.49	25.78
PATHADJACENT	AI	18.33	13.65	18.57	13.91	18.28	18.63	26.77	25.03	26.17
MUR	AI	21.89	14.56	23.44	13.71	17.84	18.12	27.98	27.86	26.24
AMUR	AI	19.67	21.16	24.34	17.75	18.13	18.00	28.00	28.00	26.31
AMUR	Human, novice	12.75	20.25	24.00	14.00	15.75	21.25	25.25	27.25	27.75
AMUR	Human, experienced	14.33	19.67	23.67	14.33	18.67	25.33	28.00	27.33	27.00
DIFFSLOW	Human, experienced	16.17	22.00	26.17	19.50	25.33	25.33	27.67	27.83	27.50

Table 1: Mean scores for each strategy and player type (AI, Human) on each map. Maps are grouped by map type.

Stated differently, the strategy difference leads to more than double the difference in scores for the PersistentGap maps than for the NoGap maps.

AI: RANDOM **vs.** PATHADJACENT **vs.** MUR **vs.** AMUR**.** Human subjects often used space as a proxy for time. A question of interest is how effective this time-space conflation is. To test this, we implemented as AI agents the strategies RANDOM, PATHADJACENT, MUR and AMUR.

RANDOM places towers at random positions. They are not guaranteed to be near a path. PATHADJACENT places towers at random spots along the path. MUR places the attack towers where they cover the largest amount of path, then does the same with freezing towers. AMUR places the attack towers where they cover the largest amount of path, then places the freezing towers at the path adjacent positions closest to the attack towers. This is a mixture of spatial and temporal strategies and was the simplest and most common strategy observed in humans.

A 4 (AI Strategy: RANDOM, PATHADJACENT, MUR, AMUR) $\times$ 2 (Map Type: NoGap, PersistentGap) $\times$ 3 (Map Number: 1, 2, 3) mixed factor ANOVA with map type and map number as within-subjects factors and strategy as a between-subjects factor performed on the scores revealed a highly significant main effect of AI strategy, $F(3, 396) = 1422.74$, $p < .001$, $MSE = 8.70$. MUR ($M = 18.26$) significantly outperformed RANDOM ($M = 9.58$, $p < .001$) and PATHADJACENT ($M = 16.90$, $p < .001$). It underperformed AMUR ($M = 19.84$, $p < .001$).

Human AMUR **vs. AI** AMUR**.** Part of building an effective spatiotemporal reasoning agent is verifying the performance of the lower-level systems. Since all strategies depend at least in part on maximizing usable range, we tested the AI implementation of this feature against novice and experienced players using the same strategy.

A 2 (Map Type: NoGap, PersistentGap) $\times$ 3 (Map Number: 1, 2, 3) $\times$ 2 (Human AMUR, AI AMUR) ANOVA revealed that the AI AMUR ($M = 19.84$) achieved a slightly higher score than the humans ($M = 18.57$), with this difference being statistically significant, $F(1, 105) = 23.35$, $p < .001$, $MSE = 2.71$. It is our belief that the modest but consistent score gap in executing this strategy likely reflects constraints on attention. When asked in interviews, why they did not select positions that covered slightly more path area subjects often responded that they had not noticed the better position but would have used it had they seen it.

Human DIFFSLOW **vs. AI** AMUR**.** To determine how the the human temporal strategy DIFFSLOW compared to predominantly spatial AI AMUR strategy, a 2 (Map Type: NoGap, PersistentGap) $\times$ 3 (Map Number: 1, 2, 3) $\times$ 2 (Strategy Type: Human DIFFSLOW, AI AMUR) ANOVA revealed a significant effect of strategy type, $F(1, 104) = 74.18$, $p < .001$, $MSE = 3.04$, with the temporal human strategy ($M = 22.42$) outperforming the spatial AI strategy ($M = 19.84$). There was also a significant map type $\times$ strategy type interaction, $F(1, 104) = 103.27$, $p < .001$, $MSE = 2.68$, with humans ($M = 21.44$) and the AI ($M = 21.72$) apparently performing similarly for the NoGap maps but humans ($M = 23.39$) outperforming AI ($M = 17.96$) on PersistentGap maps. The lack of difference in the NoGap maps can be partially attributed to particularly poor performance of two human subjects on one of the maps.

Discussion

In introducing a version of TD to study spatiotemporal reasoning, we first needed to demonstrate that the tasks selected were tractable but showed the benefit of experience. All subjects did well, with novices, on average, earning 75% of the available points across a variety of problems and difficulty levels. Experts, on average, scored 12% higher than novices. The benefit of experience varied by type of problem, with experts scoring 30% higher on PersistentGap maps.

Choice of strategy is a significant predictor of score. Novice players used spatial strategies exclusively while many (but not all) experienced players used temporal strategies. Temporal strategy users, on average, scored 13% higher than spatial strategy users overall and 30% higher on PersistentGap maps.

Much of the difference in map difficulty can be captured by a few structural properties. Scores on DynamicGap maps were 38% higher than other types of maps, independent of strategy or player experience level. It is important to understand that the higher scores are not necessarily caused by dynamic path synchronization gap. The size of the path synchronization gap grows, shrinks and reverses throughout the map due to geometry. This typically happens because the maps have complementary corners, with the line that was on the outside of a left turn becoming inside on a right turn. Complementary corners frequently cause paths to return to a location, giving towers a second chance to attack (Figure 3).

Although using space as a proxy for time has been reported in other domains, we believe this is the first time that the effects of treating space and time as equivalent on complex spatial reasoning have been assessed and modeled. The purely spatial AI MUR strategy scored 91% higher than RANDOM, indicating that re-representing temporal problems as spatial problems is highly beneficial. They are not, however, optimal. The related AMUR strategy, which supplements MUR with the temporal SLOWINRANGE strategy, scored 9% higher. Strategies that treat time and space as independent and non-proportional, potentially sacrificing spatial quantities (i.e., usable range) to increase temporal ones (i.e., tower activation time), can do much better. DIFFSLOW outperforms MUR by 23% and AMUR by 13%.

Benefits to Level Designers

The work presented here holds several possibilities for level designers and others who design and analyze maps and spatial arrangements for domains requiring strategic reasoning. Level designers must create maps of different difficulty levels to train and entertain players. Procedural content generation systems (Yannakakis and Togelius 2011) auto-generate levels, sometimes in response to player abilities and interests. Automated AI "directors" (Snowdon and Oikonomou 2011) customize a game, including difficulty level, at run time to make the game more enjoyable.

We showed that we can identify structural properties of a map that are correlated with map difficulty. This approach might be useful as part of an AI director for run-time selection of maps of desired difficulty. We have also shown that structural properties are correlated with the success of individual strategies. Combined with a recognition of the strategies a player has used, an AI director could choose maps that require a strategy the director feels the player should learn. Such an approach might also be useful in partially or fully auto-generating such maps.

The least popular maps had several structural features in common, notably a lack of switch backs (paths that pass a location more than once). We suspect that a formal analysis of the relationship between player enjoyment and map structural features would help level designers at least partially automate the design of enjoyable maps.

Conclusions and Future Work

We have identified strategies used by humans when playing TD and have shown how some of those strategies can be used by AI agents with good results. We are currently implementing the remaining strategies. We are also studying the conditions under which these strategies are most effective and testing their scalability to harder problems. All positional strategies rely in part on maximizing the amount of usable range but many are a compromise of usable range and other factors (e.g., path synchronization gap). Determining the right combination was easily done by human subjects but optimally balancing temporal and spatial features in our AI agents remains an open problem for future studies.

References

Anderson, J. R.; Bothell, D.; Byrne, M. D.; Douglass, S.; Lebiere, C.; and Qin, Y. 2004. An integrated theory of the mind. *Psychological Review* 111:1036–1060.

Arentze, T. A.; Dellaert, B. G. C.; and Timmermans, H. J. P. 2008. Modeling and measuring individuals' mental representations of complex spatio-temporal decision problems. *Environment and Behavior* 40:843–869.

Boroditsky, L. 2000. Metaphoric structuring: Understanding time through spatial metaphors. *Cognition* 75(1):1–28.

Buro, M. 2003. Real-time strategy games: a new AI research challenge. In *Int'l Joint Conf. on Artifical Intelligence*, 1534–1535.

Casasanto, D., and Boroditsky, L. 2008. Time in the mind: Using space to think about time. *Cognition* 106:579–593.

Friedman, S. E., and Forbus, K. D. 2009. Learning naive physics models and misconceptions. In *Proc. 31st Annual Conf. of the Cognitive Science Society*.

Gonzalez, C.; Vanyukov, P.; and Martin, M. K. 2005. The use of microworlds to study dynamic decision making. *Computers in Human Behavior* 21:273–286.

Jahn, G., and Knauff, M. 1998. Mental models in spatial reasoning. In Freksa, C., ed., *Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*. Springer-Verlag. 267–291.

Johnson-Laird, P. N. 1983. *Mental models*. Harvard University Press, Cambridge.

Johnson-Laird, P. N. 2001. Mental models and deduction. *Trends in Cognitive Sciences* 5(10):434–442.

Kuipers, B. J. 1978. Modeling spatial knowledge. *Cognitive Science* 2:129–153.

Lewis, C. H. 1982. Using the thinking aloud method in cognitive interface design. Technical Report IBM RC-9265, IBM, Yorktown Heights, NY.

Mandler, J. M. 2012. On the spatial foundations of the conceptual system and its enrichment. *Cognitive Science* 36(3):421–451.

Miles, L. K.; Betka, E.; Pendry, L. F.; and Macrae, C. N. 2010. Mapping temporal constructs: actions reveal that time is a place. *Quarterly Journal of Experimental Psychology* 63(11):2113–2119.

Ragni, M., and Brüssow, S. 2011. Human spatial relational reasoning: processing demands, representations, and cognitive model. In *Proc. Nat'l Conf. on Artificial Intelligence (AAAI)*, 828–833.

Schaeken, W.; Johnson-Laird, P.; and d'Ydewalle, G. 1996. Mental models and temporal reasoning. *Cognition* 60:205–234.

Snowdon, J., and Oikonomou, A. 2011. Creating more entertaining and re-playable games by dynamically introducing and manipulating, static gameplay elements and events. In *CGAMES*, 94–100.

Vandierendonck, A., and Vooght, G. D. 1997. Working memory constraints on linear reasoning with spatial and temporal contents. *Quarterly Journal of Experimental Psychology* 50A:803–820.

Yannakakis, G. N., and Togelius, J. 2011. Experience-driven procedural content generation. *IEEE Trans. on Affective Computing* 2(3):147–161.