Proceedings of the First
Artificial Intelligence and Interactive
Digital Entertainment Conference

Negative Behavior Space in the Design of Interactive Agents

Bill Tomlinson

Assistant Professor of Informatics and Drama
ACE (Arts Computation Engineering) Program, University of California, Irvine
430A Computer Science Building / Irvine, CA 92697-3425
wmt@uci.edu

Abstract

When defining the behavioral repertoire of an interactive artificial intelligence, it is helpful for designers to have a systematic way to specify which behaviors the AI will be able to perform and which behaviors it will not. The idea of a behavior space is often used in designing and implementing the behaviors that the AI will be able to perform. This paper proposes the idea of a "negative behavior space" – a set of behaviors that will explicitly not be implemented – as a useful concept for the development of AIs for games and other interactive media. Examples of negative behavior spaces are given from the development processes of two interactive installations: AlphaWolf, which was exhibited in the Emerging Technologies program at SIGGRAPH 2001, and the Virtual Raft Project, which will be shown in the Interactivity program at CHI 2005 and in the Emerging Technologies program at SIGGRAPH 2005. By explicitly excluding certain sets of behaviors from the development of an AI, designers may create interactive experiences that are clearer and more enjoyable for the player, and may also make their own work easier.

Introduction

During the process of creating real-time autonomous characters for a computer game or other interactive experience, designers must decide on the set of behaviors that each character will be able to perform. This set of behaviors is often called the character's behavior space. The behavior space encompasses all possible actions that the character may perform.

When people play computer and video games, they bring many different expectations about how the characters will behave (i.e., the contents of their behavior spaces). However, the characters will not necessarily satisfy all of these expectations. For example, a naïve player might expect the characters in one of the Final Fantasy games to be able to speak to them in the same way a real person would. However, satisfying this expectation would entail full natural language processing, comprehensive AI and expressive real-time animation at the level of a real person, all of which are well beyond the current capabilities of both game developers and academic researchers. Given the capabilities of current technologies and the goals of the game designers, it is neither possible nor desirable to satisfy every expectation that a player brings to the game.

This paper presents a concept called "negative behavior space" that can be useful in the process of defining the behavioral scope of an interactive AI. A negative behavior space is a set of behaviors that an entity is explicitly not able to perform. Working with both positive behavior spaces and negative behavior spaces can allow designers to be explicit about the implemented capabilities and intentionally designed limitations of their characters.

Related Work

The idea of negative behavior space is derived from an interdisciplinary approach to building interactive animated agents. This approach draws on ideas from computer science, visual arts, and other bodies of work.

The most relevant area of related work involves the idea of a behavior space. Goldberg and Mataric have defined the behavior space of an entity as "its set of behaviors" (Goldberg & Mataric, 1999). Mataric has also offered that "[t]he behavior space of a behavior-based system is defined by the different action sequences commanded by the various modules in the system" (Mataric´, 1992). Mali and Mukerjee define a behavior space as "a set of distinct behavior modules (i.e. no two modules have the same stimulus and consequence)" (Mali & Mukerjee, 1998). Artificial intelligence and autonomous agent researchers often use the idea of a behavior space to represent the scope of all behaviors that an entity has the ability to perform, e.g., (Lester & Stone, 1997; Marsella et al., 2000; Wray & Laird, 2003). The idea of a behavior space is often used, either explicitly or implicitly, in the development of autonomous characters, e.g., (Burleson et al., 2004). Additionally, researchers have utilized hierarchical behavior spaces (Durfee & Montgomery, 1991), behavior spaces treated as physical spaces (Killeen, 1992), behavior spaces for robotics (Goldberg & Mataric, 1999), and a coherence-structured behavior space framework for sequencing the behaviors of animated pedagogical agents (Stone & Lester, 1996). Behavior spaces are useful both as a conceptual tool in the design of behavioral entities, and also as part of the action selection process for those entities.

The second main area from which this work is the idea of "negative space," one of the core principles in graphic design (Wong, 1993), drawing (Lewis, 1984), painting (Robertson & McDaniel, 1999), sculpture (Zelanski, 1995), screenwriting (Cooper, 1994), comic book art (McCloud, 1993/2003) and numerous other areas of the arts. This technique involves attending to the space *around* objects rather than on the objects themselves. This technique helps the draftsperson to represent reality more accurately because their drawing technique is not skewed by their innate biases about how things ought to look. For example, when drawing a face, the draftsperson might have the expectation that the head is essentially oval, that the eyes are about halfway up, and that the ears are on the sides. These preconceptions could bias his or her perception and execution of the drawing, and might lead to an inaccurate final product. If, on the other hand, the draftsperson were to draw the spaces around the face, the subconscious biases would not come into play, and the drawing might be portrayed more accurately. By focusing on the negative space, rather than (or as well as) the positive space, the draftsperson may create an image that more effectively mirrors reality. As well as being useful for representing positive object, negative spaces may also be used as compositional elements in their own right.

The combination of behavior space from AI with negative space from the visual arts leads to the idea of negative behavior space, which is the topic of this paper. Negative behavior space is a concept that helps to specify the scope and limits of an entity's behavior. In this regard, it is related to other mechanisms for scoping and limiting problems, from the constraints built into computational systems (Williams & Hogg, 1994) to the entire process of design (Simon, 1996). Requirements engineers, too, use both positive and negative requirements (Bergman & Mark, 2002). Negative behavior space can be a useful part of the process of player expectation management in interactive games.

Understanding the expectations of participants is also of relevance to this paper. Bailey presented an approach to automatic story generation that focused on the role of the reader in the process (Bailey, 1999). Mateas offered that viewer expectations could be leveraged to enhance the believability of characters (Mateas, 2002). Rewards of various kinds are used systematically in computer and video games to reinforce and satisfy players' expectations (Pagulayan et al., (In press)). An awareness of the expectations that people bring to an interactive experience can allow a designer to choose to satisfy or intentionally violate those expectations, rather than accidentally violating them. As Johnson points out, failing to meet a participant's expectations can have unintended negative effects (Johnson, 2003).

Negative Behavior Space Concept

A negative behavior space is a set a behaviors that is explicitly excluded from the behavioral repertoire of an entity. Negative behavior spaces only make sense in the context of positive behavior spaces (which, for brevity, will simply be referred to as "behavior spaces"). A negative behavior space carves away some subset of the behaviors that might otherwise be included in the behavioral repertoire of an entity.

The purpose of a negative behavior space is to reduce the amount of unintentional expectation violation that occurs when a player interacts with an autonomous system. Expectation violation can be a useful technique for creating compelling games; for example, having a non-player character jump out from behind a wall is a form of expectation violation that can generate surprise and positive excitement. However, unintentional expectation violation, such as when the game pad suddenly stops working, can be annoying and exasperating for a player. Programming autonomous behavior provides an unfortunately rich set of opportunities to produce unintentional expectation violations that have the same effect on the player as a broken game pad. For example, autonomous characters sometimes get stuck in navigational traps, where their navigation algorithm is unable to cope with a certain location. A balance needs to be struck between what is expected by the player and what is desired and implementable by the production team.

The idea of a negative behavior space is a common part of many development processes, but is rarely made explicit. Failures at effectively defining the negative behavior spaces of an entity are common, and result in products that are considered to be "too busy," "distracting," "noisy," "misleading," or "full of red herrings." Negative behavior spaces could be used to help understand these kinds of problems more explicitly, and therefore to reduce their prevalence in interactions with autonomous systems.

Examples

This section details two examples from previous work in which the negative behavior space concept was a useful tool in designing virtual characters.

AlphaWolf

The AlphaWolf installation (see Fig. 1) was an interactive experience in which people could interact with the members of an animated wolf pack (see Fig. 2). Each of three participants played the role of a newborn pup in the virtual pack, and could direct the behavior of that pup by howling, growling, whining and barking into a microphone. The goal of the project was to demonstrate research in computational social relationships (Tomlinson & Blumberg, 2002) and interactions with animated characters (Tomlinson et al., 2002) by allowing people to interact with the virtual wolves. AlphaWolf premiered in the Emerging Technologies program at SIGGRAPH 2001. For more information on this project, please view a

Figure 1: Two virtual wolf pups play in the AlphaWolf installation.

segment from the television show Scientific American Frontiers with Alan Alda (PBS, 2002).

In creating this installation, the goal was to focus people's attention on the social relationships formed by the virtual wolves. This focus led to the wolves having a well fleshed out behavior space in the area of social relationships. To determine which behaviors were appropriate and necessary, background research was conducted on real wolves, e.g., (Mech, 1999). Real wolves exhibit many different kinds of social behaviors – dominance behaviors such as growling and standing with a raised tail, submissive behaviors such as whining and rolling over on the ground, play behaviors such as bowing and chasing, howling and other visible and audible actions.

In addition, a subset of reproductive behaviors such as courtship and mating are closely related to the other social behaviors. However, due to factors relating to social

Figure 2: Several participants interact with AlphaWolf at Ars Electronica 2002.

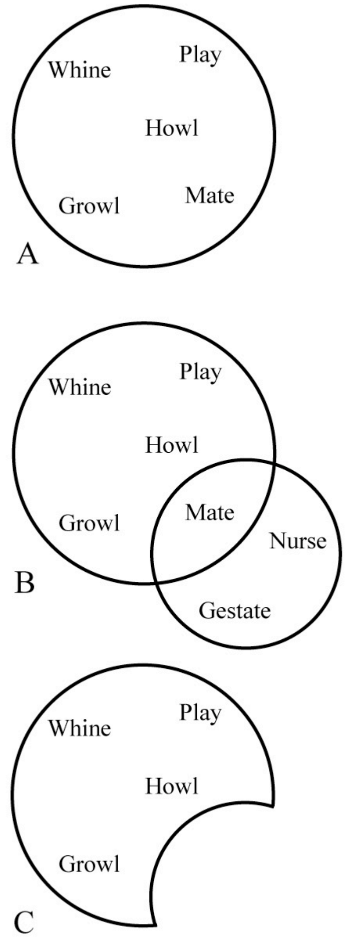

Figure 3: A demonstration of the use of negative behavior space in planning the AlphaWolf installation. A) The focus of AlphaWolf was on wolf social behaviors. The circle delineates the social behavior space with several social behaviors listed explicitly. B) Implementing mating behavior would imply gestation and nursing behaviors. The second circle represents the reproductive behavior space. C) The reproductive behavior space is subtracted from the social behavior space, leaving a subset of those behaviors to be implemented.

appropriateness and breadth of appeal, as well as to limitations of development time, the decision was made to exclude reproductive behaviors in AlphaWolf. Since these two behavior spaces – social behaviors and reproductive behaviors – overlap with each other, it was decided to implement wolf social behaviors except those that dealt directly with reproduction (see Fig. 3). In this example, social behavior was a positive behavior space and reproductive behavior was a negative behavior space. The implemented set of behaviors included all of those behaviors inside the social behavior space and outside the reproductive behavior space.

There were several other examples of negative behavior spaces that factored into the development of the virtual wolves. For example, hunting behaviors were excluded because of the difficulty of implementing a second species. The space of behaviors that wolves do with their mouths, such as carrying sticks, biting, etc., was also excluded because of the difficulty of implementing the inverse kinematics of mouth contact. The notion of a negative behavior space as an explicit concept emerged over the development process of the AlphaWolf project. Negative behavior spaces were not explicitly represented in the code of the system, but rather were part of the designers' conceptualization process.

Virtual Raft Project

A second example of a system in which the development process involved the negative behavior space concept is the Virtual Raft Project (see Fig. 4). This project will be shown in the Interactivity program at CHI 2005 and in the Emerging Technologies program at SIGGRAPH 2005. The goal of this project is to enable animated characters to break free from the constraints of a single graphical display, and jump seamlessly from those displays onto the screens of mobile devices. In the installation, several desktop screens serve as islands of virtual space, inhabited by small populations of animated virtual characters. Participants interact with this system by means of "virtual rafts" – Tablet PCs with animated rafts on them. When a participant tilts one of the Tablet PCs, accelerometers in the device detect its motion, and the raft slides down the screen, leaving graphical ripples in its wake. When the participant brings the device up to one of the virtual islands, a character can jump from the desktop screen to the Tablet PC, landing on the virtual raft. The character then has to balance on the raft as it slides around on the Tablet PC's screen. Once a character has jumped onto the Tablet PC, the participant can carry that character to another island. Once the raft is close enough to this other island, the character will jump off the raft. Several virtual islands and several virtual rafts enable multiple participants to interact with the installation at the same time. The

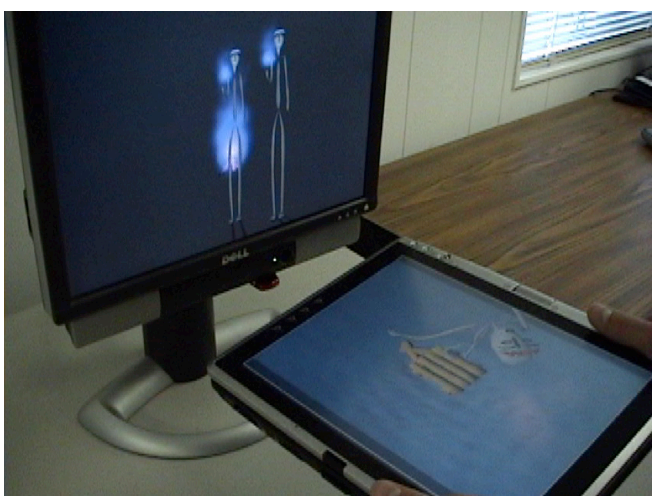

Figure 4: When the virtual raft is close enough to a virtual island, characters can jump on or off.

Figure 5: A child reaches out his hand, expecting one of the characters to jump onto it.

"Island Metaphor" of this project offers a rationale for the relationship of real space and virtual space. For more information on the Virtual Raft Project, please watch the following short online video (Tomlinson, 2004).

In this project, the audience is meant to focus on the characters' ability to engage in heterogeneous animation – animation that occurs across different computational platforms. This focus requires that the characters have three main behavior spaces – the set of behaviors that occur on the virtual island, the set that occur on the virtual raft, and the set that carries them between the two platforms. In defining each set of behaviors, both positive and negative behavior spaces were employed.

For example, on each Tablet PC, the screen is covered by simulated water on which the virtual raft floated. Because of the presence of the water, there is the expectation that the character may be able to fall in. For the implementers, having the character fall in the water was not particularly difficult, but it implied several other behaviors. A character that can fall in the water should be able to swim, climb out of the water onto the raft from different directions, appear wet once it has climbed out, and exhibit a range of other behaviors. Since these behaviors were not relevant to the core premise of the project, and/or were too difficult to program and animate, the implementers decided that water-oriented behaviors needed to be a negative behavior space, and falling in the water could therefore not be implemented.

The focus on heterogeneous animation in the Virtual Raft Project also leaves open other capabilities that participants might expect from the characters. For example, one child, upon seeing a character jump from the desktop screen to the Tablet PC, held out his hand and expected a character to jump onto it (see Fig. 5). This kind of category mistake demonstrates the broad range of possible expectations that participants bring to installations, especially those that include novel interaction technologies. Carefully defining the perimeter of a character's range of behaviors using both positive and negative behavior spaces can help avoid confusion resulting from participants' unfulfilled expectations.

In this project, as in AlphaWolf, negative behavior spaces were integral in the planning and design stages of the project. However, negative behavior spaces did not have an explicit representation in the code of the system. In future projects, it would be interesting to encode negative behaviors explicitly in the AI of the characters, thereby enabling those characters to take action to avoid the perimeters of the negative behavior spaces.

Benefits

There are two main benefits to working conceptually with negative behavior spaces. First, and most importantly, it can help to create a better experience for the players. Because there will be fewer behaviors that mislead them into expecting capabilities that AIs do not have, players will be more satisfied with those AIs. By helping to create clearer and more unified behavioral repertoires, negative behavior spaces can help to minimize unexplained limitations in character AIs (and therefore the need to explain away those limitations). In general, players are willing to accept arbitrary constraints as the "rules of the game," as long as those constraints are clear and consistent. Negative behavior spaces can help to create behavioral repertoires that have clear, intentionally designed limits. Because these limits are explicit and formally acknowledged, they may more easily be integrated into the back story, scenario or rule set that accompanies the game.

The second main benefit is that negative behavior spaces can reduce the amount of work that needs to be done by the implementers to achieve a satisfactory final product. By planning rigorously in the early stages which suites of behaviors will be implemented and which will not, fewer animations and fewer lines of code will need to be discarded. In addition, by working with both positive and negative behavior spaces (especially when adapting legacy systems that have an abundance of existing functionality), programmers may avoid giving entities functionality simply because it is possible. Often it is necessary to remove functionality from a legacy system, and negative behavior spaces can help make that process clearer.

Shortcomings

There are a number of difficulties that face interactive AI designers when attempting to define the behavior spaces of their products. First, their mental model of players' expectations is based on imperfect information – game designers do not know exactly who will play the game, or what previous experiences those players will have had. AIs must be designed based on what the designer expects from a future audience, rather than on the reality of the audience at the time of playing. An effort to define behavior spaces in advance based on the players' expectation spaces is inevitably an imperfect process. This uncertainty factors into the process regardless of whether or not the idea of negative behavior spaces is utilized, and demonstrates the utility of pilot studies and user tests to get a sense for what players actually expect.

A second challenge in crafting behavior spaces is that players' expectations are dynamic, and will change over the course of the interaction itself. Especially in a complex or long-lasting interaction, elements introduced early on can dramatically skew how later elements are perceived. An awareness of how participants' expectations will develop over the course of the interaction can help negative behavior spaces be used more effectively.

A third issue, which again applies to positive behavior spaces as much as negative behavior spaces, is that they are hard to visualize. Both forms of behavior space are frequently multidimensional and challenging to conceptualize, especially when attempting to relate them to one another. Nevertheless, the difficulty in visualizing behavior spaces does not invalidate their potential usefulness in conceptualizing behavior.

Conclusion

When people approach a computer game or other form of interactive entertainment, they have expectations about the entities that they encounter in that medium. Due to limitations in the technologies available and also in the ability to predict these expectations, interactive AIs cannot live up to these all of them. To design an interactive AI that is as satisfactory as possible, it is important for the designer to choose both the behavior spaces to populate, and the behavior spaces to avoid intentionally.

This paper has presented the idea of a negative behavior space as a useful conceptual tool for the design process of interactive entities. A negative behavior space is a set of behaviors that an entity is explicitly not capable of performing. As the famous sculptor Auguste Rodin is reputed to have said, "I choose a block of marble and chop off whatever I don't need." Similarly, negative behavior spaces can help remove portions of potential behavior space that are unnecessary, distracting or misleading to an audience. By being explicit about the sets of behaviors that will not be implemented for an AI, designers may understand their decisions more clearly, ultimately resulting in less work for the implementers and more enjoyable experiences for players.

Acknowledgements

The author would like to thank the members of Bruce Blumberg's Synthetic Characters Group at the MIT Media Lab, who collaborated on AlphaWolf, and the members of his research group at UCI, who helped to create the Virtual Raft Project. This research was supported by the ACE (Arts Computation Engineering) program, the Donald Bren

School of Information & Computer Sciences and the Claire Trevor School of the Arts at UCI, by the UC Game Culture & Technology Lab, and by the California Institute of Telecommunications & Information Technology.

References

Bailey, P. 1999. Searching for Storiness: Story-Generation from a Reader's Perspective. *AAAI Fall Symposium on Narrative Intelligence.*

Bergman, M., & Mark, G. 2002. Exploring the Relationship between Project Selection and Requirements Analysis: An Empirical Study of the New Millennium Program. *Proceedings of the IEEE Joint Requirements Engineering Conference (RE 2002),* pp. 247-254. Essen, Germany.

Burleson, W., Picard, R.W., Perlin, K., & Lippincott, J. 2004. A Platform for Affective Agent Research. *Workshop on Empathetic Agents, International Conference on Autonomous Agents and Multiagent Systems.* Columbia University, New York, NY.

Cooper, D. 1994. *Writing Great Screenplays for Film and TV*. New York, Macmillan.

Durfee, E., & Montgomery, T. 1991. Coordination as distributed search in a hierarchical behavior space. *IEEE Transactions on Systems, Man, and Cybernetics* 21(6), 1363-1378.

Goldberg, D., & Mataric, M.J. 1999. Coordinating Mobile Robot Group Behavior Using a Model of Interaction Dynamics. *Proceedings, Autonomous Agents '99*. Seattle, WA.

Johnson, W.L. 2003. Interaction tactics for socially intelligent pedagogical agents. *In IUI '03: Proceedings of the 8th int'l conference on Intelligent user interfaces*, pp. 251--253. ACM Press.

Killeen, P.R. 1992. Mechanics of the animate. *J Exp Anal Behav.* 57(3), 429-463.

Lester, J., & Stone, B. 1997. Increasing Believability in Animated Pedagogical Agents. *Proceedings of the First Int'l Conference on Autonomous Agents.*

Lewis, D. 1984. *Pencil Drawing Techniques.* Watson-Guptill Publications, New York, NY.

Mali, A., & Mukerjee, A. 1998. Metrics For Evaluation of Behavior-Based Robotic Systems. *Proceedings of ICRA (IEEE Int'l Conference on Robotics and Automation),* pp. 1122-1127. Belgium.

Marsella, S.C., Johnson, W.L., & LaBore, C. 2000. Interactive Pedagogical Drama. *Proceedings of the 4th Int'l Conference on Autonomous Agents*, pp. 301-308.

Mataric´, M.J. 1992. Behavior-Based Systems: Main Properties and Implications. *Proceedings of IEEE Int'l Conference on Robotics and Automation, Workshop on Architectures for Intelligent Control Systems*, pp. 46-54. Nice, France.

Mateas, M. 2002. *Interactive Drama, Art, and Artificial Intelligence*. PhD. Dissertation, School of Computer Science. Carnegie Mellon University, Pittsburgh, PA.

McCloud, S. 1993/2003. Time Frames (from *Understanding Comics*). In *The New Media Reader*, Wardrip-Fruin, N., & Montfort, N., Eds. MIT Press, Cambridge, MA.

Mech, L.D. 1999. Alpha status, dominance, and division of labor in wolf packs. *Canadian Journal of Zoology* 77, 1196-1203.

Pagulayan, R.J., Keeker, K., Wixon, D., Romero, R., & Fuller, T. (In press). User-centered design in games. In *Handbook for Human-Computer Interaction in Interactive Systems*, Jacko, J., & Sears, A., Eds. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

PBS. 2002. The Intimate Machine, http://pbs-saf.virage.com/cgi-bin/visearch?user=pbs-saf&template=template.html&query=Alpha+Wolf&category=0&viKeyword=Alpha+Wolf (last viewed: 4/1/05). In Scientific American Frontiers.

Robertson, J., & McDaniel, C. 1999. *Painting as a Language: Material, Technique, Form, Content.* Wadsworth Publishing, Australia.

Simon, H. 1996. *The Sciences of the Artificial*. MIT Press, Cambridge, MA.

Stone, B., & Lester, J. 1996. Dynamically sequencing an animated pedagogical agent. *Proceedings of AAAI*, pp. 424-431. Portland, OR.

Tomlinson, B. 2004. Virtual Raft Project, http://www.ics.uci.edu/~wmt/movies/VirtualRaftProject.mov (last viewed: 4/1/05). In Virtual Raft Project website.

Tomlinson, B., & Blumberg, B. 2002. Synthetic Social Relationships in Animated Virtual Characters. *From Animals to Animats 7 (SAB '02).*

Tomlinson, B., Downie, M., Berlin, M., Gray, J., Lyons, D., Cochran, J., & Blumberg, B. 2002. Leashing the AlphaWolves: Mixing User Direction with Autonomous Emotion in a Pack of Semi-Autonomous Virtual Characters. *Proceedings of the 2002 ACM SIGGRAPH Symposium on Computer Animation.* San Antonio, TX.

Williams, C., & Hogg, T. 1994. Exploiting the Deep Structure of Constraint Problems. *Artificial Intelligence* 70(73-117).

Wong, W. 1993. *Principles of Form and Design*. John Wiley & Sons, New York, NY.

Wray, R., & Laird, J. 2003. Variability in Human Behavior Modeling for Military Simulations. *Proceedings of Behavior Representation in Modeling & Simulation Conference (BRIMS).*

Zelanski, P. 1995. *Shaping Space: The Dynamics of Three-Dimensional Design*. Harcourt College Publishers, Orlando, FL.