Interactive spatiotemporal cognition:
Data, theories, architectures, and autonomy

Sangeet Khemlani and J. Gregory Trafton
{sangeet.khemlani, greg.trafton}@nrl.navy.mil
Navy Center for Applied Research in Artificial Intelligence
Naval Research Laboratory
Washington, DC 20375

Keywords: spatiotemporal cognition; interactive inference; context; pragmatics; gesture; architectures

Introduction

Everyday interactions often depend on thinking about space and time: collaborators need to know where events take place – and in what order – to, e.g., communicate driving directions, build pieces of furniture, or carry out strategic operations in military and sports settings (Núñez & Cooperrider, 2013). A simple set of driving directions may require a listener to interpret and reason about the spatial relations – such as *next to* and *behind* – and the temporal relations – such as *after* and *during* – that a speaker describes. The speaker may also use gestures to substitute, supplement, or disambiguate linguistic descriptions (Holle & Gunter, 2007; Perzanowski, Schultz, & Williams, 1998). Such rapid, rich, and productive interactions are transient and difficult to analyze behaviorally, and so they pose a challenge for experimenters. They are grounded in the physical world, and accordingly challenge computational models that cannot digest rich perceptual and environmental input in real time. Robotic systems are geared towards processing and acting upon the physical world – and they increasingly support human-robotic interaction (e.g., Fong et al., 2006; Kawamura et al., 2003; Kortenkamp et al., 1999). But they, too, are uniquely challenged in maintaining productive interactive exchanges with human teammates, because they must be tolerant of human idiosyncrasies, preferences, limitations, and errors (Trafton et al., 2013).

Because these challenges cut across broad interests in cognitive science – such as linguistics, artificial intelligence, robotics, and psychology – progress is unlikely without the engagement of multiple approaches, from psychological experimentation to the construction of autonomous, embodied systems. In recent years, progress towards understanding interactive spatiotemporal cognition has accelerated along parallel paths: there exist new behavioral and imaging methodologies to study event segmentation (e.g., Radvansky & Zacks, 2014), spatial inference (e.g., Knauff & Ragni, 2013), and gestural cognition (e.g., Novack et al., 2016); novel computational theories of understanding physical reasoning (e.g., Battaglia et al., 2013) and mental simulation (e.g., Khemlani & Johnson-Laird, 2013); cognitive architectures that support rich interactivity (Huffman & Laird, 2014; Trafton et al., 2013); and a wide variety of technological platforms on which to transform theory into embodied interaction.

The goal of the workshop is to allow these parallel approaches to converge. Discussants will share recent data and theory, consider novel architectural approaches, and demonstrate burgeoning technological advances that advance the science of spatiotemporal inference. The workshop will promote interdisciplinary collaboration by focusing on three unifying themes.

Spatiotemporal cognition and interaction

The workshop will focus on three general themes that draw from and synthesize varying approaches and methodologies in the cognitive sciences. We summarize each theme below:

The representation of time and space

People represent space and time through a variety of cognitive and neural structures (e.g., Bonato et al., 2012; Gibbon & Church, 1990; Schaeken, 1996); and computer scientists and roboticists have proposed efficient methods of extrapolating spatial and temporal structure from the external environment (e.g., Tang, Fei-Fei, Koller, 2012). This set of talks will leverage recent advances in cognitive neuroscience, computational modeling, and artificial intelligence to make progress towards efficient, domain-general representations of time and space.

Embodied communication

Gestures are often communicative (e.g., Hostetter, 2011) and can improve thinking across many tasks (e.g., Alibali et al., 2011; Bucciarelli et al., 2016; Chu & Kita, 2011; Ehrlich et al., 2006). In particular, gestures appear to be outward signs of mental simulations (Hegarty et al., 2005; Hostetter & Alibali, 2008, 2010). How do gestures help build, refine, or disambiguate mental representations? This set of talks builds upon the previous theme to understand how physical interactions depend on – and facilitate – the construction of representations of space and time.

Architecture and inference

Inferences about space and time occur in the context of building representations from environmental input and using them to act upon the world, e.g., by issuing motor commands to carry out particular gestures. Cognitive and robotic architectures allow researchers to build robust systems that connect psychological theories of spatiotemporal representation and reasoning with interactions in the physical world. This set of talks brings together researchers whose goal is to build integrated architectures for spatiotemporal reasoning systems.

Workshop structure

The workshop will include three sets of twenty-five minute presentations (including Q&A) organized around three central themes: a) representing space and time; b) gestural and embodied communication; and c) architectures for inference and interaction. Each presentation will conclude with a panel discussion. Presentations will include overviews of theoretical and computational models, discussions of recent experimental phenomena, descriptions of novel technologies, and discussions of setting an agenda for future research.

Organizers

Sangeet Khemlani is a cognitive scientist (Ph.D., Princeton University) at the Naval Research Laboratory in Washington, DC. His research focuses on the development of a unified cognitive model of deductive, probabilistic, kinematic, and temporal reasoning. His recent work focuses on integrating theories of event segmentation of temporal inference (Khemlani et al., 2015).

Greg Trafton is a cognitive scientist (Ph.D., Princeton University) at the Naval Research Laboratory in Washington, DC, who studies cognitive robotics and human robot interaction. His expertise is in building high fidelity computational cognitive models that highlight an aspect of human cognition and then putting those models on embodied platforms so that they take a person's fallacies or weaknesses into account and use that information to either compensate or facilitate the person's goals and actions.

Target audience

The workshop is targeted to a broad spectrum of researchers and practitioners in the cognitive sciences. Because it incorporates several different themes, from computational modeling, to gesture and embodied cognition, to high-level inference, it is appropriate to many disciplines within cognitive science, including psychologists, computer scientists, linguists, and roboticists. Furthermore, the workshop proposes a multidisciplinary approach towards understanding the nexus of interaction and spatiotemporal cognition.

References

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science, 22,* 1138–1144.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*, 18327-18332.

Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: Evidence for a mental time line. *Neuroscience and Biobehavioral Reviews, 36,* 2257-2273.

Bucciarelli, M., Mackiewicz, R., Khemlani, S., & Johnson-Laird, P.N. (2016). Children's creation of algorithms: simulations and gestures. *Journal of Cognitive Psychology.*

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. Journal of Experimental Psychology: General, 140, 102–116.

Ehrlich, S. B., Levine, S., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology, 42,* 1259–1268.

Fong, T., Kunz, C., Hiatt, L. M., & Bugajska, M. (2006). The human-robot interaction operating system. In *Proceedings of Human-Robot Interaction.*

Gibbon, J., & Church, R. M. (1990). Representation of time. *Cognition*, *37*, 23-54.

Hegarty, M., Mayer, S., Kriz, S., & Keehner, M. (2005). The role of gestures in mental animation. *Spatial Cognition and Computation, 5,* 333–356.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of cognitive neuroscience*, *19*, 1175-1192.

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin, 137,* 297–315.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15,* 495–514.

Hostetter, A. B., & Alibali, M. W. (2010). Language, gesture, action! A test of the gesture as simulated action framework. *Journal of Memory & Language, 63,* 245–257.

Huffman, S. B., & Laird, J. E. (2014). Learning procedures from interactive natural language instructions. In *Proc. 10th International Conference on Machine Learning. Amherst, MA* (pp. 143-150).

Kawamura, K., Nilas, P., Muguruma, K., Adams, J. A., & Zhou, C. (2003). An agent-based architecture for an adaptive human-robot interface. In *Proceedings of the 36th Hawaii International Conference on System Sciences.*

Khemlani, S., Harrison, A. M., & Trafton, J. G. (2015). Episodes, events, and models. *Frontiers in Human Neuroscience*, 9, 590, 1-13.

Khemlani, S., & Johnson-Laird, P. N. (2013). The processes of inference. *Argument & Computation*, 4, 1-20.

Kortenkamp, D., Burridge, R., Bonasso, R. P., Schreckenghost, D., & Hudson, M. B. (1999). An intelligent software architecture for semiautonomous robot control. In *Autonomy Control Software Workshop, Autonomous Agents 99* (pp. 36–43).

Novack, M., Wakefield, E. & Goldin-Meadow, S. (2016). What makes a movement a gesture? *Cognition, 146,* 339-348.

Núñez, R., & Cooperrider, K. (2013). The tangle of space and time in human cognition. *Trends in cognitive sciences*, *17*, 220-229.

Perzanowski, D., Schultz, A. C., & Adams, W. (1998). Integrating natural language and gesture in a robotics domain. In *Intelligent Control (ISIC) 1998* (pp. 247-252). IEEE.

Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition.* Oxford University Press.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review, 120,* 561-588.

Schaeken, W., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition*, *60*, 205-234.

Tang, K., Fei-Fei, L., & Koller, D. (2012, June). Learning latent temporal structure for complex event detection. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 1250-1257). IEEE.

Trafton, J. G., Hiatt, L.M., Harrison, A. M., Tamborello, F.P., II, Khemlani, S., & Schultz, A.C. (2013). ACT-R/E: An embodied cognitive architecture for human-robot interaction. *Journal of Human Robot Interaction*, 2, 30-55.