Inferring Narrative and Intention from Playground Games

Christopher Crick
Yale University
New Haven CT
Email: christopher.crick@yale.edu

Brian Scassellati
Yale University
New Haven CT
Email: scaz@cs.yale.edu

***Abstract*—We present a system which observes humans participating in various playground games and infers their goals and intentions through detecting and analyzing their spatiotemporal activity in relation to one another, and then builds a coherent narrative out of the succession of these intentional states. We show that these narratives capture a great deal of essential information about the observed social roles, types of activity and game rules by demonstrating the system's ability to correctly recognize and group together different runs of the same game, while differentiating them from other games. Furthermore, the system can use the narratives it constructs to learn and theorize about novel observations, allowing it to guess at the rules governing the games it watches. For example, after watching several different games, the system figures out on its own that Tag-like games require close physical proximity in order for the role of "it" to swap from one person to another. Thus a rich and layered trove of social, intentional and cultural information can be drawn out of extremely impoverished and low-context trajectory data.**

I. Introduction

Every schoolchild knows that teachers have eyes in the backs of their heads. From across the playground at recess, out of the corner of an eye, they always notice when a friendly game of tag transforms into something superficially similar but much more likely to end in stained clothes, torn jeans and tears. We all have this capacity; we infer a great deal about the intentions, goals, social relationships and rules of interaction from watching people interact, even from such a distance that only gross body movement can really be seen. How else could we enjoy a football game from the upper bleachers? What's more, we do not need even that much context – we are driven to recognize dramatic situations even when presented with extremely simple cues. Animated boxes on a flat white screen are enough to trigger this inference process, and we happily spin narratives of anthropomorphized black squares playing with, hiding from, bullying and swindling one another.

Making sense of very low-context motion data is an important cognitive task that we perform every day, and yet it depends on very little information from the world – so little, in fact, that we can have some hope at designing computational processes that can manipulate the manageable quantity of data to accomplish similar results. A computer or robot system able to watch groups of people perform tasks or play games and figure out their goals and the rules by which they act has a number of obvious applications: entertainment, assistive technology, monitoring and safety. In addition, we hope that the process of building systems that can perform this kind of inference with real people and in real-world situations may help illuminate and ground the psychological questions of how these capabilities develop and operate in humans.

Our system begins with positional data taken from real humans playing games with one another, derived from an installed laboratory sensor network. True, this represents a bit of a shortcut – humans have no sensory apparatus that provides us with instantaneous physical location in a 3-d coordinate space. But we do have a huge visual computational apparatus that manages to translate retinal impulses into coherent, persistent, reliable object concepts. At the current state of the art in computational perception, this is still not possible – though work like ours may help to make it so.

The system uses the raw moment-by-moment trajectory data to hypothesize about the current intentional states of the people involved, in terms of their tendency to move toward and away from one another. It then creates a narrative by stringing together sequences of these hypothesized intentional states and noticing relevant facts about the moments where the players' intentions shift. Armed with this information, the system can reliably classify and differentiate a number of playground games: Tag, Smear, Keepaway and Catch. Furthermore, from comparing the raw data to the narrative sequences it develops to describe what it saw, the system can infer some of the hidden rules of the games it sees. For instance, it works out by itself the fact that in Tag, a person can become the new "it" only when the old "it" approaches sufficiently close to tag.

II. Related Work

Our approach begins with the pioneering work of Heider and Simmel more than half a century ago [1]. They found that even simple moving geometric shapes are often perceived in animate, goal-directed terms. Humans tend to attribute goals, roles, social relationships, histories and intentions to such shapes, even when the available context is extremely impoverished. This is notable because such experiences seem to reflect automatic (and even irresistible) visual processing and are insulated to some degree from other aspects of cognition. Such percepts seem relatively unaffected by perceivers' intentions and beliefs, but are tightly controlled by subtle aspects of the displays themselves (for a review, see [2]. In addition,

such percepts seem to occur cross-culturally [3], and even in infancy [4], but they can be disrupted by neuropsychological conditions such as autism spectrum disorder [5], [6] and amygdala damage [7]. Perhaps because such phenomena seem to lie at an interesting intersection of perception and cognition, they have attracted the interest of cognitive psychologists [8], social psychologists [9], [10], developmental psychologists [11], [12], cognitive neuroscientists [13], vision researchers [14], anthropologists [3] and computer scientists [15]. A central goal of this work has been to identify the visual cues that trigger the perception of animacy such as sudden direction and speed changes [14], interactions with spatial contexts [16], apparent violations of Newtonian mechanics [17], or other characteristic movements [18].

The specific analysis undertaken by our system, hypothesizing vectors of attraction and repulsion between agents and objects in the world in order to explain the causal relationships we note in an interaction, relates to the dynamics-based model of causal representation proposed by Wolff [19] and on Talmy's theory of force dynamics [20]. Humans can explain many events and interactions by invoking a folk-physics notion of force vectors acting upon objects and agents; our system explicitly generates these systems of forces in order to make sense of the events it witnesses.

The system discussed in this paper carries forward research we reported last year [21]. That work demonstrated that an analysis based on motion trajectories could recognize events that agreed with human evaluations of the same data, an insight that the current research has leveraged into a much more sophisticated system capable of evaluating a variety of scenarios, extracting important narrative elements, making subtle comparisons and generating meaningful interpretations in an unsupervised fashion.

III. Setup

A. Detecting Motion

One of the major strengths of our approach is that the trajectory information upon which the system operates derives from real-world human interaction in an ethologically valid context. To obtain this data, we employ a system of ultrasonic beacons, sensors and triangulation software to track the locations of the game participants and objects to within a few centimeters, generating five position reports per sensor per second – fast enough to create fairly smooth motion trajectories at normal human running speeds, and to notice important events in significantly less than a second. These beacons send messages to one another using a simultaneous radio broadcast and ultrasound chirp, and the receiving units calculate distance by comparing the time difference of arrival between the two signals. The system then uses triangulation from sensors in known locations throughout the room to triangulate the position of the participants, each of whom wore a baseball cap affixed with a uniquely-identified sensor. [1]

[1]For a complete description of the sensor hardware, see [22]. The embedded control and data analysis software are our own.

The sensor system produces a five-dimensional vector of data: the identity of an object, its x, y and z coordinates in the room, and the time at which these coordinates were measured. Further details of the operation and deployment of the network can be found in [21].

B. Characterizing Attraction and Repulsion

The basic narrative building blocks employed by our system are sets of hypothesized attractions and repulsions derived from a folk-physical interpretation of the observed data, as suggested by the social force-dynamic theory described in [19] and [20].

For each agent and object in the observed environment, the system calculates the "influence" of the other people and objects on its perceived two-dimensional motion, expressed as constants in a pair of differential equations:

$$V_{x_i^n} = \frac{c_{x_j}(x_j^n - x_i^n)}{d_{ij}^n} + \frac{c_{x_k}(x_k^n - x_i^n)}{d_{ik}^n} + \cdots \quad (1)$$

(and similarly for the y dimension). We obtain the (noisy) velocities in the x and y direction, as well as the positions of the other agents and objects, directly from the sensor data:

$$V_{x_i^n} = \frac{x_i^{n+1} - x_i^n}{t_{n+1} - t_n} \quad (2)$$

(again, also for the y dimension). Here, $V_{x_i^n}$ represents the x component of agent i's velocity at time n. x_i^n, x_j^n and x_k^n are the x coordinates of agents i, j and k respectively, at time n. Likewise, d_{ij}^n and d_{ik}^n are the Euclidean distances between i and j or i and k at time n.

This results in an underconstrained set of equations; thus to solve for the constants we collect all of the data points falling within a short window of time and find a least-squares best fit. For further details, see [21]. Given the number of participants in our evaluation scenarios and the 5 Hz speed with which our sensor network furnished position reports, this window needed to be 0.5 seconds long – still a source of temporal inaccuracy, but much improved over the two seconds reported in our previous work.

C. Identifying Intentions Over Time

Each constant determined by the process described above represents in some fashion the influence of one particular agent or object on the motion of another at a particular point in time. Some of these may be spurious relationships, while others capture something essential about the motivations and intentions of the agents involved.

To determine the long-term relationships that do represent essential motivational information, we next assemble these basic building blocks – the time-stamped pairwise constants that describe instantaneous attraction and repulsion between each agent and object in the room – into a probabilistic finite state automaton, each state representing a set of intentions that extend over time. At any particular point in time, any particular agent may be attracted or repelled or remain neutral with respect to each other object and agent in the room;

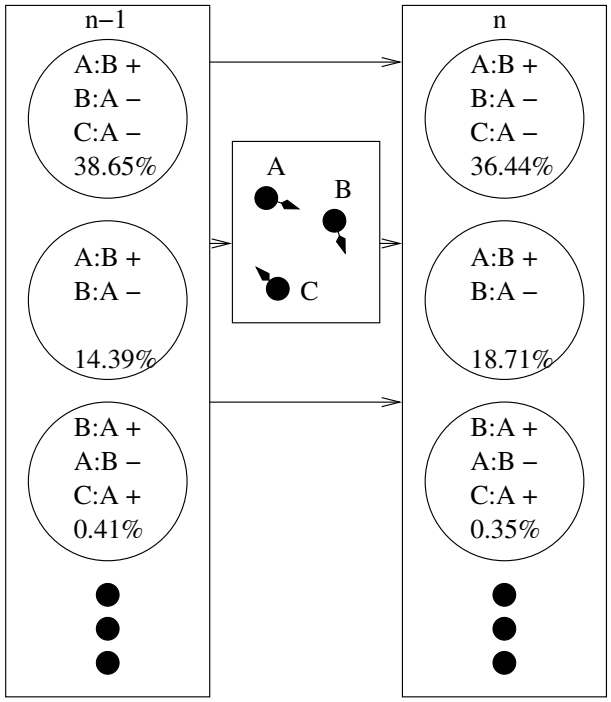

Fig. 1. The system's level of belief in a few intentional states, evolving as new information arrives. At time $n-1$, the system believes that A is intending to chase B, while B and C are fleeing from A. At time n, new data arrives that shows A continuing to chase B, but C moving sideways. Accordingly, the system's belief that B and C both want to escape from A declines (top circles), while the belief that C is neutral (middle circles) increases. More of the same kind of data would have to come in before the system would switch its belief to the middle-circle state, at which point it would review its observational history to determine the point where C actually stopped running from A. The bottom circles represent B and C chasing A, while A wants to evade B – a situation that the system currently views as very unlikely.

this is characterized by the pairwise constants found in the previous step. The system assumes that the actors in the room remain in a particular intentional *state* as long as the pattern of hypothesized attractions, repulsions and neutralities remains constant, discounting noise. A particular state, then, might be that A is attracted by B and neutral toward C, B is repelled by A and neutral toward C, and C is repelled by A and neutral toward B. This state might occur, for instance, in the game of tag when A is it and has decided to chase B.

The system maintains an evolving set of beliefs about the intentions of the people it observes, modeled as a probability distribution over all of these possible states. As new data comes in, the current belief distribution is adjusted, and the system assumes that the most likely alternative reflects the current state of the game.

$$Bel_n(S) = \frac{Bel_{n-1}(S)(1 + \lambda \sum_{c \in S} s(c_n))}{Z} \qquad (3)$$

Here, the belief in any particular state S at time n is the belief in that state at time $n-1$, modified by the current observation. c_n is the value at time n of one of the pairwise relationship constants derived from the data in the previous step; the function s is a sign function that returns 1 if the constant's sign and the intention represented by the current state agree, -1 if they disagree, and 0 if the state is neutral toward the pairwise relationship represented by the constant. λ is a "learning rate" constant which affects the tradeoff between the system's sensitivity to error and its decision-making speed. Through trial and error, we found a value of $\lambda = 0.08$ to yield the best results. Finally, Z is a normalizing constant obtained by summing the updated belief values across all of the states in the system. For an example of how the beliefs of a few of these states evolve, see Figure 1.

The function s depends on the signs, not the values, of the constants obtained in the previous step. It is perfectly straightforward to adjust the function to take account of the magnitudes of the social forces involved, but we discovered that doing so increased the effect of sensor noise unacceptably, without causing a noticeable improvement in performance.

Notice that the system never explicitly handles ambivalence in its computations – the states that accommodate neutral attitudes never update their beliefs based on the values of any constant that may exist between the neutral pair. Instead, the degree of belief in these states changes owing to normalization with other states that are changing based on attraction and repulsion. This works perfectly well, however, because it amounts to a decision to "default" to a neutral state – if our degree of belief in both pairwise attraction *and* repulsion falls, then naturally our belief in a state of ambivalence between the pair should increase.

This is not an approach which can deal easily with large numbers. The number of states in the system depends exponentially on the number of objects and agents being tracked. Each state represents a particular configuration of attractions, repulsions and neutralities between each agent pair. Although the number of states is perfectly manageable with four participants, it will eventually become too computationally complex to represent this way. However, there is no reason to believe that humans are any more capable of keeping track of many people's simultaneous motivations and goals, either. Making sense of crowd interactions may require a whole different set of assumptions, representations and computational machinery.

D. Constructing Narrative

The process described in the preceding section converts instaneous, noisy velocity vectors into sustained beliefs about the intentional situation that pertains during a particular phase of a witnessed interaction. As the action progresses, so too do the system's beliefs evolve, and as those beliefs change, the sequence of states becomes a narrative describing the scenario in progress. This narrative can be analyzed statistically to identify the action in progress, differentiate it from other possible activities, and also provide the system with clues to use in unsupervised feature detection.

When the system finds that it now believes the action it is watching has changed state, it retroactively updates its beliefs – the point at which the belief changed is *not* the actual time point at which the witnessed behavior changed. Rather, several rounds of data collection caused the change in belief, so if the

system is to reason about an actual event that changed the participants' intentional state, it must backtrack to the point where the new data began to arrive. This it accomplishes by looking at the derivative of the beliefs in the two states. The system assumes that the actual event occurred at

$$\mathbf{max}(t)(\frac{d}{dt}Bel_t(S_{\mathbf{old}}) \geq 0 \vee \frac{d}{dt}Bel_t(S_{\mathbf{new}}) \leq 0) \quad (4)$$

That is, the most recent time point when either the system's belief in the old state was not decreasing or the system's belief in its new state was not increasing.

The system can now string these states and durations into an actual (admittedly dull) narrative: "State 13 for 4.6 seconds, then state 28 for 3.9 seconds, then state 4 for 7.3 seconds..." It can translate the states into actual English: "A was chasing B for 4.6 seconds while C was running from A, then B was chasing C for 3.9 seconds while A stood by..." It can collect statistics about which states commonly follow which others (a prerequisite for developing the ability to recognize distinct activities). And it has identified points in time where important events take place, which will allow the system to notice information about the events themselves.

E. Differentiating Activity

The system has now constructed a narrative, an ordered sequence of states representing the hypothesized intentions of the people participating. These intentions ebb and flow according to the progress of the game and the shifting roles of the actors as they follow the rules of their particular game. How to evaluate this narrative to determine whether it amounts to anything we would recognize as understanding the game?

One way, tentatively explored in [21], is to compare the narrative with ones constructed by human observers of the same events. This is an attractive approach as it directly compares human and machine performance, but usually in qualitative fashion. In this work, we decide instead to use a statistical approach and determine whether the generated narrative suffices for the system to distinguish easily between different games. At the same time, it should be able to recognize that the narratives it constructs regarding different instances of the *same* game are sufficiently similar.

To do this, we treat the narratives (perhaps naturally) as linguistic constructions, and apply similarity metrics which are most often used in computational linguistics and automated language translation. In this work, we use Dice's coefficient [23], which compares how often any particular word (or in our case, intentional state) follows any other in two separate texts (or narratives).

$$s = \frac{2n_t}{n_x + n_y} \quad (5)$$

Here, n_t is the total number of bigrams (one particular state followed by another) used in both narratives, while n_x and n_y are the number in the first and second narratives, respectively.

As an example, take comparison between two games of Tag. One has 21 unique state transitions of the kind suggested in Figure 1, such as a transition from believing that C is fleeing from A to one where C is neutral toward A. Another game has 26 transitions, and the two games share 20 state transitions in common. In this case, the coefficient would be $\frac{2(20)}{21+26} = 0.86$.

This is obviously an extremely simple heuristic, and indeed Dice's coefficient does not perform as well as state-of-the-art statistical comparisons for linguistic analysis [24], but it has the advantage of simplicity and does not rely on a large narrative corpus to function.

F. Unsupervised Rule Recognition

Building step-by-step from applying force dynamics to real-word data to create hypotheses about individual intentional states, through establishing a probabilistic belief about the sequence of intentional states in a particular narrative, to comparing and contrasting narratives to find similarites and differences, our system has by now developed a fairly sophisticated and rich understanding of the social framework of the activities that it has watched. Furthermore, it possesses the tools to make its own educated guesses at contextual elements separate from the trajectory information which it used to build up its narratives. We demonstrate this by letting the system look at relative position information (*not* relative motion, which is what it has been using up to this point) and allowing it to form its own hypotheses about whether position is significant to the rules of the games it observes.

The system has already identified the moments where the intentional states of the actors change during gameplay. Why do these changes happen? Something in the rules of the game, most likely. And something to do with relative or absolute position is a good guess. Associated with each particular change in intentional state (the state bigrams discussed in section III-E), the system collects statistics on the absolute position of each actor and the relative distances between them all. If the mean of the relative distances is small, say, $\mu < 30$ cm, or, likewise, if the standard deviation of the absolute positions is small ($\sigma < 30$ cm), then it reports that it has found a potential rule, such as: in this game, and similar ones, the role of chaser and chased switch only when the two come into close contact. In other words, Tag!

IV. Experimental Validation

A. Games

To evaluate the performance of the system, we instructed test subjects to play several common playground games for periods of three minutes at a time. The sensor network tracked four objects during the period of play, games played either with four people or with three people and a ball. The system had no independent knowledge of which objects were human and which were inanimate. Six trials of each game were played, three each by two independent groups of people (undergraduate and graduate students). The games were:

- Tag: The person designated as "it" chases and tags one of the other players, at which point that player becomes "it" and the roles switch (Figure 2).

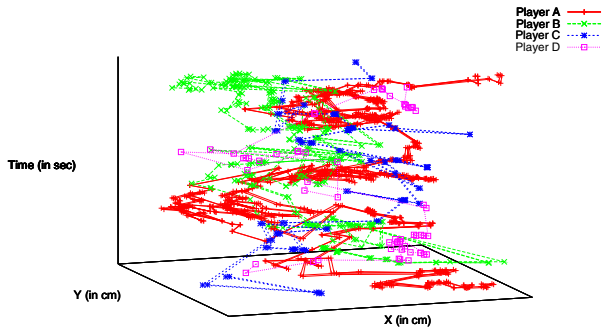

Fig. 2. A four-person game of Tag. All four people are constantly in motion, looping about the room as they chase and are chased in turn.

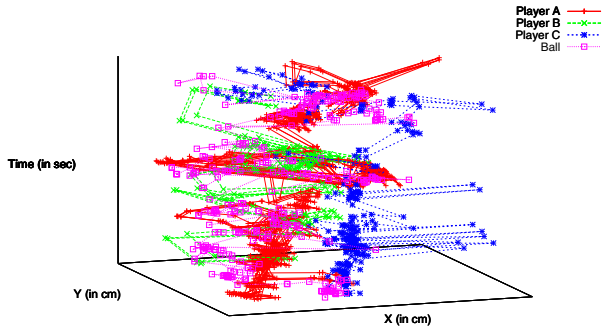

Fig. 3. A three-person game of Keepaway. The ball is constantly in motion, as is *one* of the players, while the other two players are relatively stationary. The identity of the moving player changes several times during the game.

- Smear: One of the players carries a ball while the others attempt to catch and steal the ball for themselves. Whoever carries the ball is the target of everyone else in the game.
- Keepaway: Two players throw the ball back and forth, while a third tries to intercept the passes. If successful, the interceptor becomes a thrower and the person responsible for throwing the bad pass must now attempt to intercept (Figure 3).
- Catch: Each player throws the ball to one of the other two players in turn.

TABLE I
SIMILARTY METRICS (DICE'S COEFFICIENT)

	Tag	Smear	Keepaway	Catch
Tag	0.80	0.62	0.38	0.28
Smear	0.62	0.72	0.54	0.30
Keepaway	0.38	0.54	0.66	0.66
Catch	0.28	0.30	0.66	0.86

TABLE II
CLASSIFICATION RATES

Tag	85%
Smear	77%
Keepaway	53%
Catch	83%

TABLE III
SAMPLE LOCATION STATISTICS FOR TAG

	$\mu(d_{AB})$	$\mu(d_{AC})$	$\mu(d_{BC})$
A chasing → B chasing	21.6	104.8	184.3
A chasing → C chasing	58.3	18.0	112.5
B chasing → A chasing	28.4	90.5	83.8
B chasing → C chasing	107.9	141.6	21.4
C chasing → A chasing	217.9	29.3	191.4
C chasing → B chasing	106.7	137.1	29.3

B. Activity Differentiation

For each game, we calculated the similarity metric between it and each of the other games – five other instances of the same game and six of each of the others. Table I shows the average of these intra-and inter-game similarity metrics. Tag, Smear and Catch all were significantly more similar to themselves than to any other game. The exception, Keepaway, was indistinguishable from Catch. Interestingly, the converse was not true – although an average game of Keepaway is as different from other Keepaway games as from Catch, games of Catch are much more similar to each other than to Keepaway.

Table II depicts a performance measure based on classification rates. For each game, we looked at the five games that the system judged as most similar and scored how many of those were actually the same game. If performance was perfect, then for each of the six trials of each game, the system would pick out all five of the other games of the same type, leading to 30 correct answers per game. In none of the games was the performance this good, but all are well above random (22%). Keepaway once again underperforms.

In any event, these results show that the system was indeed able to perceive the intentions of the game participants and construct narratives that captured essential similarities and differences between and among games. Given any two games, by hypothesizing about the intentions of the players and stringing together intentional states into a narrative, the system could usually tell whether the two scenarios told similar stories, and thus whether they were likely different versions of the same game or different games altogether. Interestingly, it even was able to work out that certain games, though different, were more closely related than others – Tag and Smear, which both involve chasing and tagging other people, looked more like each other than Keepaway and Catch, which both involve tossing a ball around, and vice versa.

C. Rule Recognition

The average number of narrative state transitions detected by the system across the 24 3-minute games was 28 – not a huge corpus to analyze for location-based rules. However, the system already knows which scenarios are similar enough

to be considered the same game, enlarging its pool of data significantly. For each game, the system averaged together the absolute and relative location statistics for the five most similar games, found as described in the previous section. The system identified all of the following (by displaying noteworthy state transitions and associated measurements, which the authors have translated into text for illustration):

- Tag and Smear: when the role of chaser switches between players, those two people are close together.
- Smear: the two chased objects are always next to each other (presumably the ball and the ball carrier)
- Catch: three people are always standing in the same place whenever the ball changes direction.

Table III shows an abridged example of the relative location data used to make these determinations (for clarity, only three players are shown, not four). In the six games most similar to a particular exemplar of Tag, for every type of intentional state transition, the table depicts the average distance (in cm) between the three players. In every case, the two players who are swapping roles are located close to one another.

V. Conclusion

Humankind, it has been suggested [25], should more appropriately be called *Homo narrans*: storytelling man. We are driven to tell stories to make sense of the world, even or perhaps especially when we have only sparse information to go on. We have built a computational system that begins to demonstrate how a rich set of inferences and conclusions can be drawn from rudimentary motion data, leading to largely correct judgments about activity classification and rule inference. To do this, it hypothesizes about human intentions and constructs narratives around real-world social interaction. The better the story it can tell, the more powerful the conclusions it can draw about the activities it sees.

We are working to expand the range of activity and the detail of the narrative structure available to our machines. We have demonstrated the system's ability to pick out important rules governing the interactions it witnesses, but so far those judgments are made entirely after the fact. Hopefully, the system will soon be able to incorporate this information to improve its own process of narrative construction – for example, if the system has begun to guess that touching is an important component of Tag, then it should expect touching at particular points, and that should affect the evolution of its beliefs about the state of the game. Furthermore, a rigorous treatment of narrative expectation should also help in trying to wean the system away from the abiological sensor systems we currently use. If we have good guesses on where to look next, it may make the visual object-tracking problem a little easier. Finally, we plan to integrate this system more closely with robotic participants in the environment. If there is anything people enjoy more than telling stories about other people, it is telling stories that include themselves. The same may be true of robots.

Acknowledgment

Support for this work was provided by a National Science Foundation CAREER award (#0238334) and NSF award #0534610 (Quantitative Measures of Social Response in Autism). This research was supported in part by a software grant from QNX Software Systems Ltd, and by the Sloan Foundation.

References

[1] F. Heider and M. Simmel, "An experimental study of apparent behavior," *American Journal of Psychology*, vol. 57, pp. 243–259, 1944.

[2] B. J. Scholl and P. D. Tremoulet, "Perceptual causality and animacy," *Trends in Cognitive Sciences*, vol. 4, no. 8, pp. 299–309, 2000.

[3] H. C. Barrett, P. M. Todd, G. F. Miller, and P. W. Blythe, "Accurate judgments of intention from motion cues alone: A cross-cultural study," *Evolution and Human Behavior*, vol. 26, pp. 313–331, 2005.

[4] P. Rochat, T. Striano, and R. Morgan, "Who is doing what to whom? young infants' developing sense of social causality in animated displays," *Perception*, vol. 33, pp. 355–369, 2004.

[5] A. Klin, "Attributing social meaning to ambiguous visual stimuli in higher functioning autism and asperger syndrome: The social attribution task," *Journal of Child Psychology and Psychiatry*, vol. 41, pp. 831–846, 2000.

[6] M. D. Rutherford, B. F. Pennington, and S. J. Rogers, "The perception of animacy in young children with autism," *Journal of Autism and Developmental Disorders*, vol. 36, pp. 983–992, 2006.

[7] A. S. Heberlein and R. Adolphs, "Impaired spontaneous anthropomorphizing despite intact perception and social knowledge," *Proceedings of the National Academy of Sciences*, vol. 101, pp. 7487–7491, 2004.

[8] G. Gigerenzer and P. M. Todd, *Simple heuristics that make us smart*. Oxford University Press, 1999.

[9] R. A. Mar and C. N. Macrae, "Triggering the intentional stance," in *Empathy and Fairness*, G. Block and J. Goode, Eds. Chichester, UK: John Wiley & Sons, 2006, pp. 110–132.

[10] T. Wheatley, S. C. Milleville, and A. Martin, "Understanding animate agents: Distinct roles for the social network and mirror system," *Psychological Science*, vol. 18, pp. 469–474, 2007.

[11] V. Dasser, I. Ulbaek, and D. Premack, "The perception of intention," *Science*, vol. 243, pp. 365–367, 1989.

[12] G. Gergely, Z. Nadasdy, G. Csibra, and S. Biro, "Taking the intentional stance at 12 months of age," *Cognition*, vol. 56, pp. 165–193, 1995.

[13] S.-J. Blakemore, P. Boyer, M. Pachot-Clouard, A. Meltzoff, C. Segebarth, and J. Decety, "The detection of contingency and animacy from simple animations in the human brain," *Cerebral Cortex*, vol. 13, pp. 837–844, 2003.

[14] P. D. Tremoulet and J. Feldman, "Perception of animacy from the motion of a single object," *Perception*, vol. 29, pp. 943–951, 2000.

[15] V. Gaur and B. Scassellati, "Which motion features induce the perception of animacy?" in *Proceedings of the 2006 IEEE International Conference for Development and Learning*, Bloomington, Indiana, 2006.

[16] P. D. Tremoulet and J. Feldman, "The influence of spatial context and the role of intentionality in the interpretation of animacy from motion," *Perception & Psychophysics*, vol. 68, no. 6, pp. 1047–1058, 2006.

[17] R. Gelman, F. Durgin, and L. Kaufman, *Causal cognition: A multidisciplinary debate*. Oxford: Clarendon Press, 1995, ch. Distinguishing between animates and inanimates: Not by motion alone, pp. 150–184.

[18] A. Michotte, *Feelings and emotions: The Mooseheart symposium*. New York: McGraw-Hill, 1950, ch. The emotions regarded as functional connections, pp. 114–125.

[19] P. Wolff, "Representing causation," *Journal of Experimental Psychology*, vol. 136, pp. 82–111, 2007.

[20] L. Talmy, "Force dynamics in language and cognition," *Cognitive Science*, vol. 12, pp. 49–100, 1988.

[21] C. Crick, M. Doniec, and B. Scassellati, "Who is it? inferring role and intent from agent motion," in *Proceedings of the 11th IEEE Conference on Development and Learning*, 2007.

[22] N. B. Priyantha, "The cricket indoor location system," Ph.D. dissertation, Massachusetts Institute of Technology, 2005.

[23] L. R. Dice, "Measures of the amount of ecologic association between species," *Journal of Ecology*, vol. 26, pp. 297–302, 1945.

[24] F. J. Ochs and H. Ney, "Improved statistical alignment models," in *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, 2000.

[25] J. D. Niles, *Homo narrans: the poetics and anthropology of oral literature*. University of Pennsylvania Press, 1999.