How do 100 people walk a tightrope together? An experiment in large scale joint action

Daniel C. Richardson (dcr@eyethink.org)
Cognitive, Perceptual & Brain sciences, University College London
Gower Street, London WC1E 6BT, UK

Rick Dale (radale@memphis.edu)
Department of Psychology, The University of Memphis
202 Psychology Building, Memphis, TN 38152, USA

John Rogers (john@delosis.com)
James Ireland (james@delosis.com)
Delosis, 8 Grosvenor Road, Twickenham, Middlesex TW1 4AE, United Kingdom

Abstract

A lecture hall full of people played a computer game together. Their goal was to keep a tightrope walker balanced. Each had a handset that could deliver a left or right nudge. The tightrope walker was also pelted by tomatoes which knocked him off balance. Across several games, the difficulty of the task was changed by the frequency of tomatoes and whether or not they were visible. After each game, the participants rated their own and the group's performance. We analysed the button presses of individuals, and quantified how they related to the moment by moment action of the group and movement of the tightrope walker. On successful games, participants were able to anticipate the behaviour of the group and kept the tightrope walker in equilibrium.

Keywords: joint action; wisdom of crowds; group behaviour, situated cognition

Introduction

There is wisdom and beauty in a crowd. Galton (1907) studied competitions to guess the weight of a cow, a common game at village fares. He noted that the average response of the crowd usually equalled or bettered any of the individual guesses. We now know that if the faces of all those villages were averaged too, they would beat any individual villager in a beauty contest (Langlois & Roggman, 1990). These principles have been extended into business decisions, analysing markets and predicting political events (Surowiecki, 2004). In each case, the claim is that the average of group's response is superior to individual's judgements, even when those individuals are thought to be experts. The same idea applies to a large number of judgements made by a single person: one's own average estimate is better than any single guess (Vul & Pashler, 2010). One explanation is that the biases that distort individual judgements (or facial characteristics) are roughly randomly distributed. Polling a large number of people or decisions evens out these distortions. The principle is that if incompetence is normally distributed, then the average response will be wise.

But is the superiority of crowds restricted to wisdom? In all these cases, single judgements or measurements are being made in response to static problems or criteria. What about governing continuous action, when a stream of decisions have to be made in time, in response to changing circumstances? In short, there may be wisdom in a crowd but what happens when they have to *act* together?

Around the time of Galton, people were very interested in 'the mob', and the possibility of understanding a crowd as if it were an organism with a single mind (e.g., Le Bon, 1896, Freud, 1921). Analysis of the behaviour of large groups became the domain of sociology and political science, however, as psychology focused experimental tools on the individual. Social forces themselves are studied in social psychology of course, but perception and action are typically studied in their absence. The laboratory cubicle of a typical cognitive psychologist is a lonely place. More recently, that has been changing.

A diverse set of researchers have come to the realisation that perception, action and cognition cannot be fully understood by investigating single individuals (e.g., Barsalou, Breazeal & Smith, 2007; Robbins, & Aydede, 2009; Sebanz, et al 2006). Studies of situated cognition show that cognition 'in the wild' is intimately linked not only to representations of the external world, but also to the cognitive processes of others. For example, Hutchins (1995) observed the ways that navy navigators would distribute cognitive processes between themselves by using external tools and representations, such as maps and notations. Knoblich and Jordan (2003) gave a detailed analysis of the way that two people coordinate their actions. To be successful, participants had to anticipate both the movements of the objects in the game and the actions of their partner.

In our experiment, over a hundred people played a computer game together. Our first goal was to see if the ability of crowds to make good judgements (Surowiecki, 2004) also meant that they could successfully act together in a dynamic task. Our second goal was to take predictions about pairs of participants acting together (Knoblich & Jordan, 2003) and see if they scale up to much larger groups.

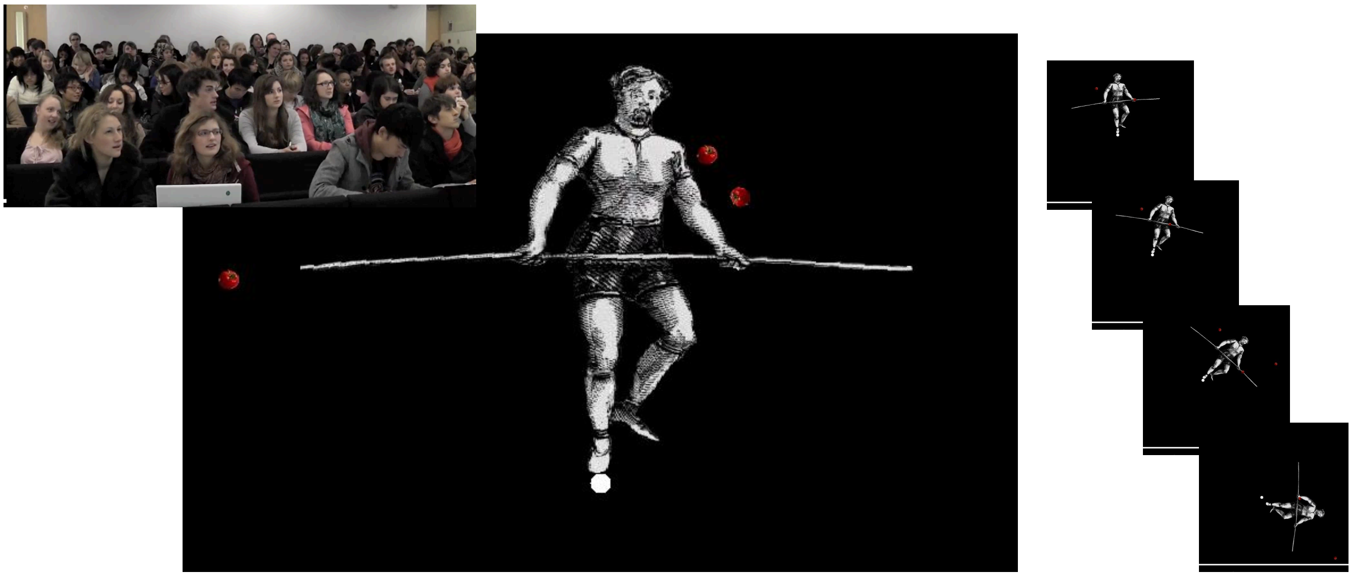

Figure 1. Screen capture from the tightrope walker game and (inset) the participants controlling him. On the right, the impact of a tomato causes the end of a game

The tightrope game

We developed a simple game that could be played by a large number of people simultaneously[1]. Participants saw on a projection screen a picture of a man holding a pole, balanced on rope (Figure 1). Each participant held a handset and pressed one of two buttons. A laptop computer collected the responses and controlled the movements of the tightrope walker. Each time one of the participants pressed a button, it immediately sent a very small nudge to the tightrope walker, sending him to the left or right. The movements of the tightrope walker were governed by a physics engine that accounted for the size and position of the figure and the pole and their momentum. A game ended when the tightrope walker fell off the rope.

The game was made harder by introducing random noise in the form of tomatoes. They were fired from the sides of the screen at random and knocked the tightrope walker to the left or right. The frequency of these missiles could be varied to change the difficulty of the game. Additionally, the tomatoes could be made invisible, so Bob's balance would be perturbed unpredictably.

We ran a series of 18 games, systematically varying the degree and visibility of the tomatoes, and whether (without their knowledge) the tightrope walker was being controlled by all of the participants or only half of them. All button pressed were recorded for analysis. We quantified the success of each game in terms of its duration and the tightrope walkers average deviation from the vertical, and polled participants on their view of their individual performance and that of the group as a whole. This allowed us to investigate participants' perception and evaluation of their own actions under different levels of difficulty, and develop models of how they performed the task and responded to each other.

Models: Agent Policies and Bob's Survival

Knoblich and Jordan (2003) studied the dynamics of a simple game of coordination. Pairs of participants saw a target dot move repeatedly across a screen. The participants task was to move a ring shape so that it hovered over the target. Although the target immediately reversed its direction when it reached the edges of the screen, the ring could only be sped up or slowed down in increments, each time one of the participants pressed a key for or against the current direction of motion. An optimal strategy was to anticipate when the ring would need to change direction, and begin pressing the key in the opposite direction before the turn had to be made. When one participant could use both keys, this strategy was followed. When two participants acted together, each using a different key, they had difficulty performing the task. However, if they could hear a bleep each time that their partner pressed a key, then they had little difficultly learning the strategy of anticipatory control, and performing the task to the level of an individual acting alone.

We developed a simple group dynamics model that would explore whether strategies like anticipation, and response diversity (see below), can assist the group in sustaining Bob's position on the tightrope. To do this, we simply defined a vector of button states, with as many elements as we had participants in the classroom:

$$\mathbf{v}(t) = <a_1, ..., a_{120}>$$

Each of these "agents," ai, can take on values 1, -1, or 0, depending on whether they are moving Bob to the right, left, or inactive, respectively. To initiate a simulation, we take Bob's position as being an iterated function of the current state of v, and the previous state of Bob:

[1] If this paper is presented as a talk at the Cognitive Science conference, then the audience will of course be invited to play the game

$$\mathrm{Bob}(t) = \mathrm{Bob}(t\text{-}1) + \alpha\sum\mathbf{v}(t) + \tau\alpha\sum\mathbf{v}(t\text{-}1) + N$$

So Bob's current position is a function of his last position, displaced by the summed response vector of the classroom (α is a multiplier to set how much each button press acts on Bob's position). We also included a momentum term which displaces Bob by a small proportion (τ) of the previous response, and a Gaussian noise signal (N). This is a simplification of the physics in the real game, but suffices as a test of different strategies. The model is a simple linear accumulator that will fluctuate between negative values and positive values depending on the policies that the agents (a_i) use to issue a pulse. Take the simplest policy: press left (a_i = -1) when Bob is right (Bob(t) > 0), vice versa for pressing right (a_i = +1), and otherwise do nothing (a_i = 0). The result is an oscillating Bob position.

By setting an arbitrary threshold for Bob's demise, the run of Bob(t) values will end at some point that his absolute position exceeds that threshold and he falls off the rope. We can run simulations of different agent policies to see what works best to keep Bob aloft in this context. We considered the strategy sets described below. These include two primary kinds of response strategies agents ("participants") could engage in: diversity of responding and anticipation.

Bare: For each cycle, a random 30 agents press a button in the opposite direction of current position

Anticipate: In addition to the above, a random 20 agents per cycle press a button in Bob's direction of movement *if* he is not yet past the 0 mark (in other words, press left when Bob is just about to cross over to the right)

Diversity: In addition to "bare," 20 random agents per cycle randomly select their buttons.

Both: Use the anticipate and diversity strategies together.

We ran 1,000 simulated "trials" for each, and calculated the number of cycles for which these different strategies kept Bob balancing. As can be seen in Figure 2, the strategies improve the performance of the group. Creating individual-based policies that mix different responding strategies among the group -- avoiding simplistic uniformity of responding patterns -- helps maintain the Bob(t) variable before it reaches its threshold. These two notions will be our focus in analysis of the large-scale human experiment: uniformity (or, conversely, diversity) of responding patterns, and anticipatory tendencies.

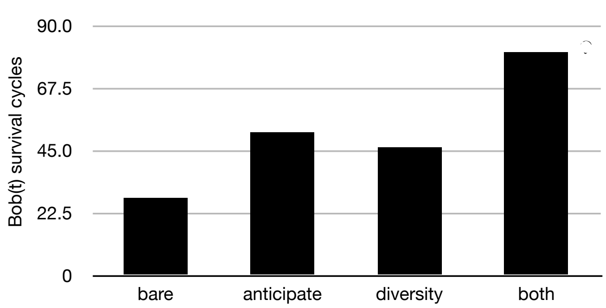

Figure 2. Game Simulations with different strategies

Methods

Participants

123 people participated in the experiment as part of a first year psychology laboratory class.

Apparatus

Each participant had a 12 button TurningPoint audience response system handset (Turning technologies).The handsets broadcast a RF signal containing the key press and handset identity number. These signals were detected by a USB receiver inserted in a Mac laptop. The game was written in Java by Delosis, and read data from the receiver using the Turningpoint API. Survey data was collected using the TurningPoint AnyWhere application. The game and survey questions were displayed on an 2x3m projection screen in front of the participants.

Design

The participants played 18 games across two blocks. In the first block, the tomatoes were not visible, and in the second they were. Within each block we randomised and counterbalanced the frequency of the tomatoes (none, low, high) with the controllers that were active (all, handsets with odd identity numbers, handsets with even identity numbers).

The experimenter gave a countdown before each game commenced. It continued until the tightrope walker fell off the rope, or until 30 seconds had passed.

After each game, four questions were displayed on screen, and participants gave their responses by pressing a button between 1 and 9:

Q1. How much control over Bob did you feel that you had?
Q2. How much did you feel that you were acting as part of a coherent group?
Q3. How do you rate your performance as an individual?
Q4. How do you rate your performance as a group?

Results

Question Ratings and Trial Time

Sense of individual control. We tested how participants' sense of individual control (Q1) related to how long they kept Bob balanced. To do so, we used a linear mixed-effects model as described in Baayen, Davidson, and Bates (2008), using participant as a random factor, and predicting sense of individual control score with total trial time in seconds (reflecting how long the group was able to keep Bob on the tightrope). There was a strong and significant positive relationship between these variables, *beta* = .37, $p < .0001$. The longer Bob balanced, the more they felt they had control over the situation.

Sense of group coherence. There was a strong and significant positive relationship between sense of group coherence (Q2) and trial time, *beta* = .40, $p < .0001$. The longer they kept Bob aloft, the more they felt they acted as a group.

Sense of individual performance. We carried out the same analysis for Q3, and also found a strong positive relationship, *beta* = .31, $p < .0001$.

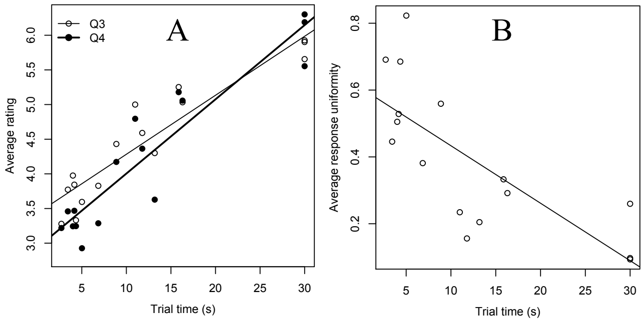

Figure 3. Correlations between trial time and average (A) performance rating, (B) response uniformity

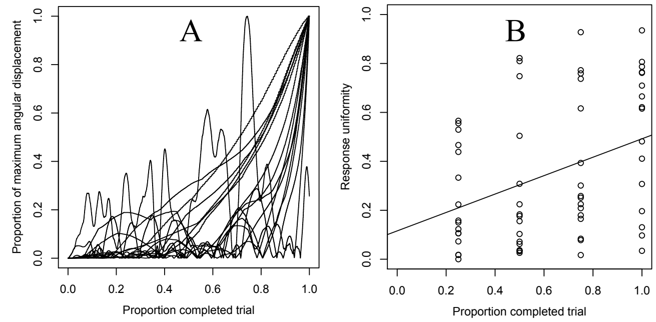

Figure 4. Proportion of trial completed, plotted against (A) proportion of max angle and (B) response uniformity

Sense of group performance. As one would expect, group performance rating (Q4) is strongly predicted by how long Bob balanced, *beta* = .40, $p < .0001$. Interestingly, as shown in Figure 3A, this relationship was different compared to Q3: The sense of individual performance was higher than sense of group performance when the group performed poorly. This is expressed as a difference in regression slopes, and can be tested by using an interaction term in a model predicting trial time. These slopes are indeed different, *beta* = .17, $p < .0001$. In other words, when the group of which participants are a member performs poorly, they are biased to attribute this gradually more to the group than to themselves.

Group Behavior

Response diversity. What predicts group success? We generated an average "uniformity score" for the group within each trial. This was based on a scoring of how uniform responses are at 5 time points with a trial (0-.25, .25-.50, etc., proportion of trial completed), then aggregated. For example, in the extreme cases, the group may generate all leftward or rightward responses, and response uniformity would be high (= 1). At the other extreme, responses may be equibiased within a trial, exhibiting maximal diversity (= 0). Uniformity tended to go down as a function of group performance (trial time), $r = -.75$, $t(14) = -4.3$, $p < .001$. This relationship is shown in Figure 3B. The greater the response diversity, on average, the longer trials lasted.

Bob's oscillatory amplitude increases. What predicts that the group will fail? First, inspection of Bob's absolute angle suggests that he is gradually being pushed to more and more extreme angular displacements as a trial proceeds. In Fig.4A, we show the time course of each trial superimposed by using proportion time instead of raw time. This figure shows that across trials, angle is gradually going up. The trial-by-trial average correlation between time and angle magnitude is .72, $t(15) = 12.2$, $p < .0001$.

Within-trial uniformity increases. It appears that one possible reason for increasing angle magnitude is that all trials involve an increase of response uniformity. This would gradually cause increasing magnitude of sway back and forth as the group's uniform responding adds momentum to Bob's sway, causing him to fall. Fig. 4B shows the score of response uniformity, as defined above, as it changes within a trial, showing a rise over time. The trial-by-trial average correlation between time and uniformity is .56, $t(15) = 6.6$, $p < .0001$. So, while successful trials sustain response diversity for longer periods of time (as shown above), within-trial failure appears to correlate with a rising uniformity in displacement of Bob.

Individual Behavior

Because the data track the responses at an individual level, it is possible to analyze the strategies employed by different participants. The following analyses are more qualitative in nature, and are meant to reflect the emergence of different agent policies that we demonstrated in the simple model used above.

Do participants engage in diversity of responding? The uniformity analyses above suggest that diversity is crucial for the group to succeed, lest Bob gain too much angular momentum as he fluctuates. Does this diversity reside at the group or individual level? In the latter case, this would suggest that participants may be gaining a sense of the collective dynamics of the group itself.

Response uniformity increased from the first to the final quarter of the trials to the end (slopes > 0, $p < .01$). But the measures at the beginning of the trials are surprisingly high, indicating that some trials involve early responding that is completely consistent within a participant. Figure 5A shows that the success of a trial (in seconds) is a function of response diversity *within subjects*, at the start of the trials, $r = -.77$, $t(14) = -4.5$, $p < .0005$.

Do participants make anticipation responses? How do individual participants respond to angle displacement? Like the simulations above, if participants responded simply, they would just wait for Bob to be leftward leaning then click the right button, and vice versa for rightward leaning. However, if participants gain a sense of the group's overall behaviour, and the need to control Bob's momentum in that context, they may choose to respond leftward just before Bob leans right, in order to control Bob. This would suggest that more sophisticated policies are emerging.

We defined an "anticipation response" as one that is made when the response is in the direction of the current angle, but opposed to the ongoing angular change. For example, if angular movement is rapidly moving rightward when Bob is leaning left, then participants who anticipate that once Bob's angle becomes right leaning, it will be difficult for the group (if responding with uniformity) to pull him back. As discussed above, this is likely the source of Bob's demise:

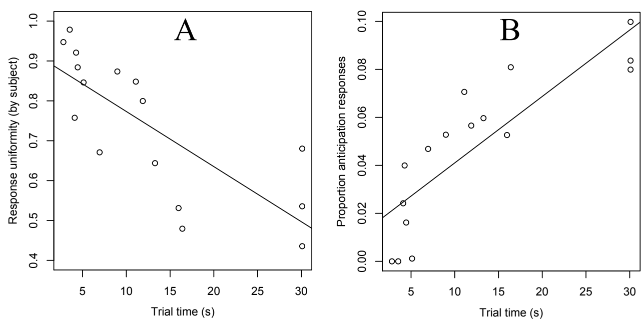

Figure 5. Correlation between trial time and (A) response uniformity (by subjects), (B) anticipations

growing uniformity of responses causes angular momentum to increase monotonically and thus Bob's angle to gradually grow until he falls. Anticipation may come from subjects, like the agents in the simulation, who wish to pull Bob back from the brink *before* the rest of the group pushes his momentum to much in the opposing direction.

Based on this logic, anticipation is occurring when the left button is pushed when Bob is in a left (negative) angle but moving rightward (positive change); vice versa, when a participant presses a right button when Bob is at a positive angle, but moving left (moving in a negative, leftward direction). The number of anticipatory responses made has a strong relationship to group performance, $r = .99$, $t(14) = 30.3$, $p < .0001$. This is a curiously orderly pattern, so we carried out a control analysis looking at the *proportion* of responses that are anticipatory. This controls for length of the trial, which may simply relate to anticipatory responses because there are proportionally more responses overall that are made. Even with this control analysis (proportion anticipation), there remains a robust relationship with group performance, $r = .84$, $t(14) = 6.00$, $p < .0001$. Thus, participants are making anticipatory responses, and the more they do so (proportionally) the more likely Bob will stay aloft (see Figure 5B).

General Discussion

What is the difference between thinking and acting alone, and thinking and acting in a social context? From the margins of traditional cognitive science, a diverse set of results have established three broad conclusions about the effect of social context on behaviour. First, as a result of social interaction, people often become more alike. Second, behaving more alike has cognitive and affiliate consequences. Third, the cognitive processes of an individual flexibly and eagerly couple with those of others: these are mechanisms of joint action (Gallantucci & Sebanz, 2009).

When two people meet, they become more like each other. They implicitly imitate each others' accent, speech rate and syntax; they look at the same things and use the same words; they adopt similar postures, gesture alike and gently sway together (Chartrand & van Baaren, 2009). Pairs of participants completing a puzzle task (Shockley et al 2003), and mobile phone users separated by miles (Murray-Smith et al., 2007) will synchronise their body movements.

Such behavioural coordination has an effect on its participants. From simply tapping in time (Hove & Risen, 2009), to copying mannerisms (e.g. Chartrand & Bargh, 1999) to aligning postures (Maurer and Tindall, 1983), mimicry can increase rapport, liking, empathy and affiliation (Chartrand & van Baaren, 2009), and how well conversants remember their interaction (Macrae et al 2008). Alignment at the level of word and syntax choice is argued to ease linguistic processing for speakers and listeners (Pickering and Garrod, 2004).

Experimental methods are starting to reveal the mechanisms involved in joint activity. In a standard stimulus-response compatibility task, participants make a judgment about one stimulus property (colour) and ignore another stimulus property (location). If there is an incompatibility between the irrelevant location property and the response (left or right finger movement), then reaction times increase, as the irrelevant property activates the incompatible response representation (Simon, 1969). Sebanz, Knoblich, & Prinz (2003) took this task and split it between two people. They sat next to each other, and each person responded to one colour: in effect, each acting one of the fingers of a participant in Simon's (1969) experiment. Although each individual had only one response to execute (and hence no need to represent the incompatible response), they still showed slower responses in the incompatible trials. When performing the same single response task alone, there was no incompatibility effect. Sebanz et al (2003) concluded that, when acting jointly, participants represented their partners' actions as if they were their own.

Studies of gaze coordination tell a similar story of social interaction, joint action, and a close coupling of behaviour. We showed two people the same scene, such as a painting, and tracked their gaze while they conversed (Richardson, Dale & Kirkham, 2007). Using the same cross-recurrence analysis tools as used here, we showed that their gaze is tightly coordinated: about three seconds before and after one person is looking at something, their conversational partner is likely to be looking at the same thing. This coordination is causally linked to comprehension. When parts of the scene that the speaker looked at are flashed, dragging a listener's gaze toward them, then the listener's comprehension improves (Richardson & Dale, 2005). This coordination changes according to the conversants' common knowledge and what they believe each other can see (Richardson, Dale & Tomlinson, 2009). Even when people are not interacting with each other, there is an effect of social context. When they are performing the same task as another person, they improve their memory for shared stimuli (Shteynberg, 2010) and even shift whether they look at pleasant or unpleasant pictures (Richardson, et al *submitted*).

The traditional view of group action is that it is individual action plus an additional compensation for the behaviour others. Our results, and the findings we have reviewed here, suggest that perhaps group action comes more naturally to the brain than it does to our theories of it.

Conclusion

Reading of the 'wisdom of the crowds' for the first time, it is tempting to imagine a world devoid of experts and pundits, in which important judgements and decisions are made by a large community of people individually clueless but together wise. But in the biggest experiments in joint action - democratic elections - we tend to favour systems that sample the full electorate infrequently, and cede the day to day decision making to a few individuals. As democracy this would be unrecognisable to the ancient Greeks, who debated and voted upon each decision before them. The common argument for our modern system (an intermittently elected oligarchy) is that it delivers strong leadership that can respond decisively to dynamic situations. In Britain and the US, such arguments are used to support two party systems and first-past-the-post voting. The fear with more representational forms of democracy is that their broad spectrum of diverse opinions would produce a mire of continual compromise, and be unable to act in a timely, coherent fashion. In our experiment, however, a large group cooperated to perform a dynamic task that required continual action. More significantly, its success rested upon disagreement and diversity of responses.

Of course, there is more to politics than simply swaying to the right or left. In our simple game, there is only one goal: to keep the tightrope walker upright. The group collaborated to achieve this goal. What will happen when the group has to select sub-goals, such as whether to move to the left or right to avoid a falling rock? Here, a simple compromise between left and right would lead to failure. Before acting together, the group would have to perceive where the majority are headed, and then follow that decision. Our future research will explore the space of different games and investigate how a large group can both think and act together.

Acknowledgments

We would like to thank Günther Knoblich, Natalie Sebanz and the participants of JAM 3 for the inspiring discussions that led to these experiments, to Lee Tbaily for assistance collecting the data, and Karen Scher at Turning Technologies for technical support and encouragement.

References

Barsalou, L.W., Breazeal, C., & Smith, L.B. (2007). Cognition as coordinated non-cognition. *Cognitive Processing, 8*, 79-91.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76,* 893–910.

Chartrand, T. L., & van Baaren, R. (2009). Human mimicry. In M. P. Zanna (Ed.), *Advances in experimental social psychology.* (pp. 219-74). Burlington: Academic Press.

Freud, S. (1921). *Group Psychology and the Analysis of the Ego.* W. W. Norton & Company

Galantucci, B., & Sebanz, N. (2009). Joint action: Current perspectives. *Topics in Cognitive Science, 1,* 255-259.

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition, 27(6)*, 949-961

Hutchins, E., (1995). *Cognition in the Wild.* MIT Press.

Knoblich G., Jordan S. (2003). Action coordination in individuals and groups: Learning anticipatory control. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29,* 1006–1016.

Langlois, J.H., & Roggman, L. (1990). Attractive faces are only average. *Psychological Science, 1,* 115-121

Le Bon, G. (1895) The Crowd: A study of the popular mind. Dover Publications.

Galton, F. (1907). Vox populi. *Nature, 75,* 450-451.

Macrae, C. N., Duffy, O. K., Miles, L. K., & Lawrence, J. (2008). A case of hand waving: Action synchrony and person perception. *Cognition, 109*, 152–156.

Maurer, R. E., & Tindall, J. H. (1983). Effect of postural congruence on client's perception of counselor empathy. *Journal of Counseling Psychology, 30*, 158–163.

Murray-Smith, R. D., Ramsay, A., Garrod, S., Jackson, M., & Musizza, B. (2007). Gait alignment in mobile phone conversations. In A. D.Cheok & L.Chittaro (Eds.), *Proceeding of MobileHCI 2007* (pp. 214–221). AMC International Conference Proceeding Series, 309

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioural and Brain Sciences, 27 (2)*, 169-90; discussion 190-226.

Richardson, D.C & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29*, 1045–1060.

Richardson, D.C., Dale, R., & Kirkham, N.Z. (2007). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science, 18(5)*, 407-413.

Richardson, D.C., Dale, R., & Tomlinson, J.M. (2009). Conversation, gaze coordination, and beliefs about visual context. Cognitive Science, 33(8), 1468-1482.

Robbins, P. & Aydede, M. (Eds.) (2009). *The Cambridge Handbook of Situated Cognition*. Cambridge, UK: Cambridge University Press.

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition, 88*, B11–B21.

Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 326-32.

Shteynberg, G., (2010). A silent emergence of culture: The social tuning effect. *Journal of Personality and Social Psychology,* Online Aug 16, 2010.

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology, 81*, 174–176.

Surowiecki, J. (2004). *The wisdom of crowds*. New York: Random House.

Vul, E. & Pashler, H. (2008). Measuring the crowd within: probabilistic representations within individuals. *Psychological Science, 19(7)*, 645-647.