# Event Segmentation of Agent Interactions: Comparing the Whole with Its Parts

**Bryan L. Koenig[1,2] (koenigbl@ihpc.a-star.edu.sg)**
**David Pautler[1] (pautlerd@ihpc.a-star.edu.sg)**
**Jonathan S. Herberg[1] (herbergjs@ihpc.a-star.edu.sg)**
**Kum Seong Wan[1] (kswan@ihpc.a-star.edu.sg)**
**Brian Monroe[1,2] (monroebm@ihpc.a-star.edu.sg)**
**Edwin Wirawan[1] (wirawane@ihpc.a-star.edu.sg)**

[1]Institute of High Performance Computing, A*STAR
[2]National University of Singapore, Singapore

**Abstract**

How do observers perceptually organize the events of individual agents when observing interactions among them? Do they readily perceive all events? Do they selectively perceive some events but not others? Do they see events overlooked by observers focusing on the individual agents? To explore these questions, participants viewed the Heider and Simmel (1944) animation, which shows three moving figures and elicits strong impressions of interacting agents. Participants in the default condition segmented the animation into meaningful events. Those in focus conditions did likewise, but focusing on one of the figures. Results indicate that participants in the default condition disregarded many events identified in the focus conditions, but identified only one event missed by focus-condition participants. These findings suggest that observers of interactions do not encode all events or gain additional insight by "seeing the big picture"; rather, they selectively perceive some events at the cost of overlooking others.

**Keywords:** Social perception; event segmentation; unit formation; perspective taking; movement cues; animation.

## Introduction

Much research has centered on how observers perceive and understand the actions of individual agents. In a typical experimental task, participants view an activity such as washing dishes and mark when, in their judgment, a meaningful event ends and another begins. These marked time points are referred to as *breakpoints*, and participants substantially agree about their placement (e.g., Zacks, Tversky, & Iyer, 2001). Breakpoint placement is not an arbitrary consequence of the segmentation task. Brain activity selectively occurs during mere observation at the same time points that participants identify as breakpoints during a subsequent segmentation task (Zacks, Swallow, Vettel, & McAvoy, 2006). These findings suggest that event segmentation reflects the perceptual structure of events.

Other research on how observers perceptually segment the events of animations of *multiple* interacting figures finds that motion cues are strongly correlated with breakpoints (Hard, Tversky, & Lang, 2006). Indeed, simple computational models can use seven motion cues to identify agentic motives in animations of two moving figures with accuracy levels similar to those of human observers (Blythe, Todd, & Miller, 1999). Absolute and relational motion cues can be quite complicated in such stimuli, which sometimes have three figures moving in relation to one another and the background context. We wondered how human observers incorporate information from each agent when viewing stimuli of multiple agents interacting. In order to address this question, we had participants in a *default condition* segment into meaningful events a multiple-figure animation previously shown to elicit compelling perceptions of social interactions (Heider & Simmel, 1944). Other participants in *focus conditions* segmented the same animation but only for events meaningful for the figure specific to the condition. We compared participants' segmentation patterns across conditions to test whether observers by default process all events relevant for all agents. Alternatively, perhaps by default observers systematically miss events that are important for some agents or notice events that would be overlooked when focusing on any one agent. We now consider evidence that suggests each of these is plausible.

Research in which participants segment perceived events for animations of moving figures suggests that movement features such as sudden acceleration elicit the perception of event boundaries (Hard, et al., 2006; Zacks, 2004). For such stimuli, participants even indicate breakpoints at similar time points when viewing multiple-agent animations both forward and backward, and these time points strongly correlate with objective movement features (Hard, et al., 2006). Such findings provide support for the notion that observers incorporate all movement cues when perceiving events, which we label the *objective movement hypothesis*. According to this hypothesis, observers perceive ongoing interactions in terms of all movement

cues from all figures. This would lead to the prediction that participants in the default condition should generate a segmentation that includes all, and no more, of the breakpoints identified across all focus conditions.

Zacks and Tversky (2001) note that for both event segmentation and object perception observers must organize parts into wholes. Heider (1958) suggested that an analogous process – unit formation – occurs during social perception wherein multiple agents, such as those interacting, are perceived as a perceptual unit. If observers perceptually group agents, the action of one agent (e.g., hitting) might be subsumed as a lower-level component of a larger schematic interaction (e.g., fighting), and thereby be disregarded. The *unit formation hypothesis* thus postulates that when observers parse ongoing events in terms of interactions they sometimes neglect agents' individual actions. This hypothesis predicts that participants in the default condition, when viewing a highly salient interaction, will sometimes miss events identified by participants in the focus conditions when the events pertain to a figure that *is part of* that interaction.

Heider (1958) also suggested that unit formation might result in observers attending to the interaction unit as the focal point of perception (the "figure"), with agents outside the interaction receiving little attention (the "ground"). The *distraction hypothesis* therefore postulates that observers attending to an interaction will sometimes fail to notice events meaningful to agents who are not part of the interaction. This hypothesis predicts that participants in the default condition, when viewing a highly salient interaction, will sometimes miss events identified in the focus conditions when the events pertain to a figure that *is not part of* that interaction.

Other research has shown that perspective-taking can interfere with an objective evaluation of events. For example, in one study people from each side of a conflict perceived a media report of an event as biased against their own side (Vallone, Ross, & Lepper, 1985). Similarly, fans of each team in an important football game, when viewing identical tapes of the game, reported events that diverged from each other (Hastorf & Cantril, 1954). This research is complemented by experiments suggesting that adults sometimes have difficulty taking another's perspective even though all of the relevant information is available to them (Keysar, Lin, & Barr, 2003). To our knowledge, perspective-taking has not been evaluated using animated figures. Nonetheless, we embody these ideas in the *perspective-taking hypothesis*, which postulates that focusing on the perspective of one figure interferes with the overall perception of events. It seems plausible that events noted by participants in the default condition but not in any focus condition could be the result of difficulty participants experience as they attempt to focus on or take the perspective of one figure. This hypothesis predicts that events perceived by participants in the default condition sometimes will be missed by participants in all of the focus conditions.

## The Current Study

We designed a study to test these four hypotheses. In it participants segmented the Heider and Simmel animation (1944) into meaningful events. The animation shows a house-like rectangle and three moving figures: a large triangle, a small triangle, and a circle (*T, t,* & *c*). Observers tend to describe the animation as a bully attacking two innocent passersby (Heider & Simmel, 1944). By comparing the default segmentation with those provided for focal agents, we evaluate how observers normally view the animation: whether perception of the whole is equal to, greater than, or less than the sum of its parts.

### Method

**Participants** Participants were 74 female and 51 male American workers on Amazon Mechanical Turk, a crowdsourcing marketplace service, who received US$ 0.90 for participating. Their mean age was 32.50 years (*SD* = 11.47).

**Stimulus Animations** For the experimental task we modified a version of the Heider and Simmel (1944) animation downloaded on 3 September 2010 from Carnegie Mellon University at http://anthropomorphism.org/img/Heider_Flash.swf. We increased its frames per second rate from 10 to 30 to make it smoother, keeping it 74 seconds long. For the practice task we created a Flash version of the hide-and-seek animation used in Hard, et al. (2006). It is 84 seconds long and depicts two squares and a circle moving as if they were playing hide and seek in their environment, which has wall-like lines. Both animations were prepared in Adobe Flash and were monochrome, 550 pixels wide by 400 pixels high, but viewed dimensions depended on participants' monitor displays.

**Design** Participants were randomly assigned to one of four conditions, which differed only in their segmentation instructions. In the *default condition* (n = 39) participants segmented the animation with no instructions to focus on any specific figure. Thus, their segmentations were potentially based on events for all of the figures. In each of the three focus conditions, the instructions were to segment the animation with a focus on the events for a specific figure, that is, for the big triangle (n = 30), the little triangle (n = 20), or the circle (n = 29).

**Procedure** Participants completed the study over the internet. They provided consent and demographic information and then read instructions indicating that (a) they would see two short animations depicting geometric figures in motion, (b) while watching the animation they should press the spacebar whenever one meaningful event ended and another began, and (c) they would briefly describe each animation after watching it. The instructions also displayed pictures of the figures. Participants then watched and concurrently segmented the practice animation (i.e., hide and seek). Once it ended, participants described what happened in each one-second interval bin for which they had pressed the spacebar. While providing descriptions, participants could review the animation but they could not add or remove markers. The instructions clarified that participants should try to provide their initial impression of the animation (at the time of the spacebar press). Once participants submitted descriptions for each marker, they could continue to the experimental task.

The procedure for the experimental task was the same as that for the practice task, except for different segmentation instructions in the focus conditions. All participants viewed the Heider and Simmel (1944) animation and instructions that indicated that the animation has three moving geometric figures, a large triangle, a small triangle, and a circle, with pictures of all figures. In the default condition, the segmentation instructions were the same as for the practice animation. The circle-focus condition included the following additional instructions: "However, this time do so only for the circle. That is, press the spacebar whenever, for the circle, a meaningful event ends and another begins." Similar instructions were added to the focus conditions for the big triangle and the small triangle. A blue arrow also pointed at the picture of the focal figure. Participants segmented the animation according to their condition's instructions, provided descriptions as in the practice task, logged their MTurk ID into the system, and were debriefed.

## Results

If a participant pressed the spacebar during a one-second bin we counted that bin as containing a breakpoint (for similar approaches see, e.g., Hard, et al., 2006; Massad, Hubbard, & Newtson, 1979). Since only *T* is visible at the start and end of the experiment animation, all results exclude the first 4 and last 12 bins, leaving 58 bins for analysis. We also excluded the data of five participants who indicated one or zero breakpoints while doing the experimental task and that of two who indicated no breakpoints in the practice task.

**Overall Segmentation** We evaluated whether focusing on one figure affected the number of one-second bins participants marked as containing a breakpoint. The objective movement hypothesis predicts that the default condition should contain the breakpoints present in all three focus conditions. If this is the case, participants in the default condition should indicate more breakpoints than participants in the focus conditions (assuming segmentation patterns differed across focus conditions, and correlations shown below indicate they did). A one-way between-subjects ANOVA on the number of breakpoints did not indicate a difference across conditions ($M$ = 8.96, $SD$ = 5.11), $F(3, 114) = 1.05$, $p = .373$, $\eta_p^2 = .03$. This lack of a difference cannot be explained by a ceiling effect; participants on average indicated only 15.45% of bins as including breakpoints. Instead, these results suggest, counter to the objective movement hypothesis, that participants in the default condition did not segment based on the objective movements of all three figures. Rather, these results suggest that participants in the default condition may have neglected some events noticed in the focus conditions, consistent with the unit formation and distraction hypotheses. Because focus-condition participants may have noticed some events missed in the default condition, but missed others noticed in the default condition, these results provide no clear evidence regarding the perspective-taking hypothesis.

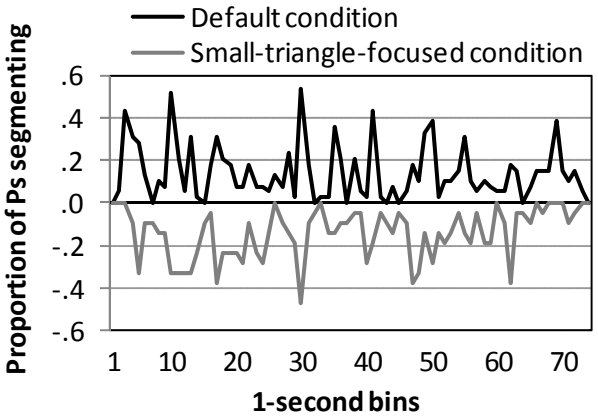


Figure 1. Proportions of participants who selected each one-second interval bin as containing a breakpoint in the default and *t*-focused conditions. The proportions for the *t*-focused condition are shown as negative to facilitate visual comparison.

**Agreement about Event Timing** To gain further insight into how participants perceptually integrate information of individual agents when perceiving them interact, we compared across conditions which one-second interval bins participants tended to indicate as containing breakpoints. For these analyses, we calculated for each 58 one-second bin the proportion of participants in each condition who indicated that the bin contained a breakpoint. Figure 1 presents the segmentation histogram for the default condition and, for comparison, the *t*-focused condition. We correlated the segmentation histogram of the default

Table 1. Correlation coefficients across conditions among segmentation histograms (proportions of participants who indicated a breakpoint in each one-second bin).

| Condition | Default | Big T. | Small T. | Circle |
|---|---|---|---|---|
| Default | 1 | .61* | .48* | .53* |
| Big T. | - | 1 | .41* | .22 |
| Small T. | - | - | 1 | .38* |

* $p < .01$. Note: T. = triangle.

condition with that of each focus condition (see Table 1). All three Pearson correlation coefficients were positive and significant; however, all fell short of the lower bound of the 95% confidence interval of the estimated correlation coefficient for participants within the default condition (estimated using bootstrap aggregation, i.e., sampling 2000 groups of n = 20 from the 39 default condition participants then calculating 58 bin means for each group and pairing groups to calculate 1000 correlation coefficients across bins; mean $r = .78$, median $r = .79$, 95% CI: .62, .90). This range was estimated to specify the approximate optimal correlation that we could expect between the default histogram and that of any other condition. The finding that the focus condition histograms were less than optimally correlated with the default condition histogram suggests that our manipulation was successful in that participants in the focus conditions were not simply responding to events involving any figure, but instead when focusing on one figure they perceived events differently.

The differences in magnitude across the three correlation coefficients were not significant (maximum $z = 1.14$, $p = .254$, using the method suggested by Meng, Rosenthal, & Rubin, 1992), but their relative magnitudes suggest that default-condition participants' segmentation may have been influenced most by *T*, then *c*, and the least by *t*. To further evaluate this possibility, we did a multiple regression analysis wherein we predicted the segmentation histogram from the default condition by the segmentation histograms from all three focus conditions. The results suggest that *T* provided a large unique contribution to the perception of the animation, $b = .47$, $p < .001$, 95% CI = .27, .67, *c* had a medium sized unique contribution, $b = .31$, $p < .001$, 95% CI = .15, .47, but *t* did not contribute any information above and beyond that provided by *T* and *c*, $b = .16$, $p = .194$, 95% CI = -.08, .40 (the intercept did not differ significantly from zero, $b = -.01$, $p = .746$, 95% CI = -.05, .04). This regression's multiple correlation coefficient (multiple-$R = .74$, $F(3,57) = 22.10$, $p < .001$) fell within the 95% CI for the estimated correlation coefficient within the default condition, suggesting that this regression performed near optimally. This suggests that participants in the focus conditions indicated many if not most of the breakpoints identified in the default condition; that is, these analyses provided no support for the prediction of the perspective-taking hypothesis that events would be missed if focusing on a specific figure. We note that exploratory regression analyses including all two-way interactions and the three-way interaction indicated no additional significant predictors, *p*s > .226.

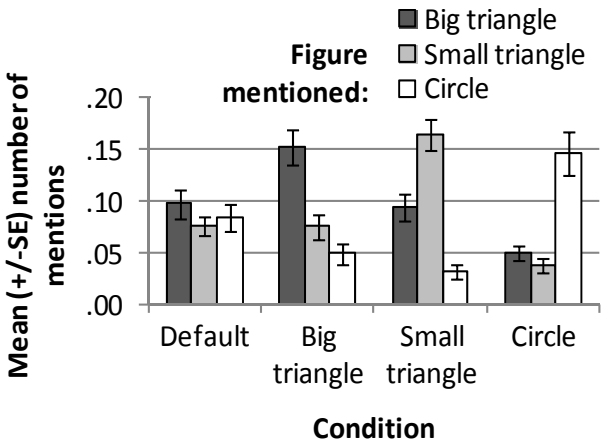


Figure 2. Mean (+/- SE) number of times each participant mentioned the figures, by condition.

**Analyses of Breakpoint Descriptions** We coded which figures participants mentioned in their breakpoint descriptions. As Figure 2 shows, participants in the focus conditions mentioned their focal figure significantly more than participants in any other condition, minimum $F(3,114) > 10.84$, $p < .001$, $\eta_p^2 = .22$. This suggests that our manipulation was successful. Participants in the default condition did not mention the three figures an equal number of times, $F(2, 76) = 8.32$, $p = .001$, $\eta_p^2 = .18$. Uncorrected post hoc tests indicated that default-condition participants mentioned *T* more than *t*, $t(38) = 5.86$, $p < .001$; they mentioned *T* marginally more than *c*, $t(38) = 2.02$, $p = .051$; but they mentioned *c* and *t* about equally, $t(38) = 0.16$, $p = .124$. These findings suggest that *T* was the most salient figure for the default-condition participants.

**Disagreement across Conditions** We also calculated for each bin a difference score equal to the proportion of participants in the default condition identifying the bin as containing a breakpoint minus the maximum such proportion among the three focus conditions (see Figure 3). Most strikingly, nearly all difference scores were negative. Indeed, a binomial test ($p = .005$) indicated fewer positive scores than expected by chance given a .25 probability that each of the four conditions would have the largest proportion. Our earlier analysis indicated no significant difference in the number of breakpoints across conditions, whereas the current finding indicates that, when comparing bin-by-bin, for almost all bins participants in one of the focus conditions were more likely to indicate a breakpoint

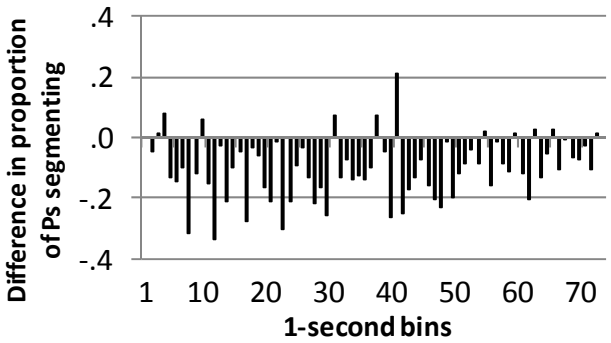


Figure 3. Difference in proportion in the number of participants indicating a breakpoint for each one-second interval bin in the default condition with the maximum proportion doing so from any of the focus conditions.

than participants in the default condition. This suggests that for observers in the default condition some events had reduced salience as compared to observers who focused on a particular figure, consistent with the unit formation and distraction hypotheses. By contrast, almost all events that are salient in the default condition remain so in the focus condition for at least one of the figures, contradicting the perspective-taking hypothesis. Indeed, setting a .2 difference in proportions as a cutoff to identify clear-cut differences in event salience, only one bin has a more salient event in the default condition as compared to any of the focus conditions. By contrast, 15 bins had events that were more salient in at least one of the focus conditions as compared to the default condition (see Table 2). Corroborating our previous findings that for default-condition participants the figures were of unequal salience, the three figures were not equally likely to be part of a reduced-salience event: *T* was in 3 such events, *c* in 5, and *t* in 8, multinomial test, $p = .017$.

Why did these 15 events have reduced salience for participants in the default condition? To determine the likely source of the reduced salience for each of these events we examined which figures were mentioned in participants' event descriptions across conditions. For convenience, we use the term "overlooked figure" to refer to the figure whose event had reduced salience in the default condition. Let us first consider which figures are mentioned in descriptions from the focus condition for the overlooked figure. If they predominantly mentioned the overlooked figure and one other figure, this suggests that the overlooked figure appeared to interact with the other figure and therefore that the reduced salience was due to unit formation. If these descriptions instead predominantly mentioned only the overlooked figure, but descriptions from the other focus conditions did not mention the overlooked figure but did mention the other two figures, this suggests that the other figures had a salient interaction that distracted default participants from noticing the

Table 2. Reduced-salience events, that is, events for which the discrepancy in the proportion of participants who indicated a breakpoint was greater by at least .2 in any focus condition (boldface) as compared to the default condition. The table reports the proportions by condition, the coded explanation for the difference (source: dist. = distraction, unit. = unit formation), and the ongoing overall event.

| | Proportion P. by condition | | | | | |
|---|---|---|---|---|---|---|
| Bin | Def. | *T* | *t* | *c* | Source | Ongoing overall event |
| 8 | .10 | **.42** | .14 | .03 | dist. | *t* & *c* arrive |
| 12 | .05 | **.39** | **.33** | .07 | unit. | *t* & *T* start fight |
| 14 | .03 | .10 | **.24** | .03 | unit. | *t* & *T* fight |
| 17 | .31 | .39 | .38 | **.59** | dist. | *t* & *T* fight |
| 21 | .08 | .13 | **.29** | .08 | unit. | *t* & *T* fight |
| 23 | .08 | .16 | .24 | **.38** | dist. | *t* & *T* fight |
| 24 | .08 | .16 | **.29** | .10 | unit. | *t* & *T* fight |
| 28 | .23 | .10 | .14 | **.45** | dist. | *t* & *T* fight |
| 30 | .54 | .23 | .48 | **.79** | dist. | *t* & *T* fight |
| 40 | .03 | .13 | **.29** | .00 | dist. | *c* & *T* fight |
| 42 | .03 | .13 | .05 | **.28** | unit. | *c* & *T* fight |
| 47 | .18 | .19 | **.38** | .31 | both | *t* joins *c* & *T* |
| 48 | .10 | .19 | **.33** | .21 | both | *t* joins *c* & *T* |
| 50 | .38 | **.58** | .29 | .28 | unit. | *T* fights *c* & *t* |
| 62 | .18 | .06 | **.38** | .28 | unit. | *t* & *c* leave |

Note: Def. = Default condition

overlooked figure. However, if descriptions from the focus condition for the overlooked figure predominantly mentioned the overlooked figure, but the other conditions had very few descriptions, we looked to recent bins for clarification. If descriptions from recent bins across conditions suggested a two-figure interaction, we coded the reduced salience as due to unit formation if the overlooked figure was part of that interaction or as due to distraction otherwise. Finally, two reduced-salience events had characteristics suggesting both unit formation and distraction. That is, descriptions from the default condition and the focus condition for the overlooked figure predominantly mentioned the overlooked figure and one other figure, suggesting the overlooked figure was interacting with the other figure and was thereby perceived as a unit with it. On the other hand, descriptions from the other two focus conditions predominantly mentioned both the other two figures but not the overlooked figure, suggesting their interaction distracted default condition participants from seeing the overlooked figure. Using these criteria, from 58 bins we clearly associated 7 with unit formation, 6 with distraction, and 2 with characteristics of both. These findings provide fairly direct support for the unit formation hypothesis and the distraction hypothesis.

## General Discussion

We investigated how observers perceptually organize the events that are meaningful for individual agents while observing them interacting with one another. Our results support the notion that observers selectively perceive some events and not others. Moreover, our results suggest that this sometimes occurs because observers link individual agents into larger perceptual units with the consequence that the salience of events at the unit-level sometimes dominates the salience of events for agents within the unit. At other times observers appear to have missed events important for one agent because their attention was focused on more salient interactions between other agents. These findings provide some empirical support for Heider's (1958) idea of unit formation. They also replicate the findings of Massad, et al. (1979) that observers selectively perceive some events and disregard others, although in their study selective perception resulted from pre-information about what would happen in the Heider and Simmel animation. Previous research has noted the importance of movement features for event segmentation, but our findings suggest that observers do not normally perceive interaction events in terms of all movement features for all agents. This finding suggests a caveat to the idea that observers use all motion information. Future research will be required to more fully determine when and why observers fail to incorporate some motion features. Our results suggest that participants were able to focus on one figure as instructed, and that doing so resulted in little difficulty due to perspective-taking. Indeed, to the extent to which participants in the focus conditions engaged in perspective-taking, our findings suggest this is not always a difficult process (cf. Keysar, et al., 2003), but can occur in a rather effortless fashion. In fact, the results indicate a greater cost, in terms of missed events, of engaging in the default rather than an agent-focused perspective.

The current study compared how observers segment events when focusing holistically on all agents to how they do so when focusing on individual agents. A future study could more completely separate information regarding single agents from the whole by removing all other figures from the animation, leaving a single, isolated figure. We could then compare segmentations of the animated solo figures with those of default and focus conditions. This would allow us to evaluate the relative importance for perceptual segmentation of context-free movement cues versus relational movement cues. Such a comparison might provide a more direct test of the unit formation hypothesis (Heider, 1958). We also note that participants in the default condition may have perceived events that they failed to report, but why they would do so more than participants in the focus conditions is unclear. We leave such problems and further questions about the perception of interactions to future research.

## Acknowledgments

We thank Barbara Tversky and Bridgette Hard for providing their hide and seek animation. We thank Barbara Tversky and three anonymous reviewers for insightful comments on earlier drafts of this paper. We also thank Swati Gupta and William Chandra Tjhi for assistance with statistical analyses.

## References

Blythe, P. W., Todd, P. M., & Miller, G. F. (1999). How motion reveals intention: Categorizing social interactions. In G. Gigerenzer, P. M. Todd, and the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 257–286). New York: Oxford University Press.

Hard, B. M., Tversky, B. & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition*, *34,* 1221–1235.

Hastorf, A. H., & Cantril, H. (1954). They saw a game: A case study. *The Journal of Abnormal Psychology, 49*, 129–134.

Heider. F. (1958). The psychology of interpersonal relations. Wiley: New York.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57*, 243–249.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind usage in adults. *Cognition, 89*, 25–41.

Massad, C. M., Hubbard, M., & Newtson, D. (1979). Selective perception of events. *Journal of Experimental Social Psychology, 15,* 513–532.

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111,* 172–175.

Vallone, R. P., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology, 49*, 577–585.

Zacks, J. M. (2004). Using movement and intention to understand simple events. *Cognitive Science, 28,* 979–1008.

Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neural correlates of event perception. *Brain Research, 1076*, 150–62.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127,* 3–21.