



Social event segmentation

Julia Boggia & Jelena Ristic

To cite this article: Julia Boggia & Jelena Ristic (2015) Social event segmentation, The Quarterly Journal of Experimental Psychology, 68:4, 731-744, DOI: [10.1080/17470218.2014.964738](https://doi.org/10.1080/17470218.2014.964738)

To link to this article: <http://dx.doi.org/10.1080/17470218.2014.964738>



Accepted author version posted online: 15 Sep 2014.
Published online: 07 Nov 2014.



Submit your article to this journal [↗](#)



Article views: 137



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Social event segmentation

Julia Boggia¹ and Jelena Ristic²

¹Department of Psychology, University of Ottawa, Ottawa, ON, Canada

²Department of Psychology, McGill University, Montreal, QC, Canada

Humans are experts in understanding social environments. What perceptual and cognitive processes enable such competent evaluation of social information? Here we show that environmental content is grouped into units of “social perception”, which are formed automatically based on the attentional priority given to social information conveyed by eyes and faces. When asked to segment a clip showing a typical daily scenario, participants were remarkably consistent in identifying the boundaries of social events. Moreover, at those social event boundaries, participants’ eye movements were reliably directed to actors’ eyes and faces. Participants’ indices of attention measured during the initial passive viewing, reflecting natural social behaviour, also showed a remarkable correspondence with overt social segmentation behaviour, reflecting the underlying perceptual organization. Together, these data show that dynamic information is automatically organized into meaningful social events on an ongoing basis, strongly suggesting that the natural comprehension of social content in daily life might fundamentally depend on this underlying grouping process.

Keywords: Event segmentation; Social attention; Social behaviour; Attention.

Humans possess a sophisticated nonverbal social communication system that has evolved to support seamless and often automatic interpretation of a range of social signals—from quick computation of where others are looking (Kobayashi & Kohshima, 1997) to evaluating the potential for social interaction, mate preference, and emotional engagement (Emery, 2000). What enables understanding of social information during everyday dynamic situations? One intriguing possibility is that environmental content is organized into units of “social” perception (i.e., social events), which in turn guide social understanding and behaviour. We tested this idea using the unit marking

procedure, which indexes the formation of perceptual units based on the features available in the environment (Kurby & Zacks, 2008; Newton, 1973; Newton & Engquist, 1976; Zacks, Braver, et al., 2001) while measuring participants’ attentional allocation during the task by recording their eye movements. We found that participants parsed social information differently than nonsocial information and that the formation of social event boundaries was reliably related to attention paid to the information conveyed by eyes and faces.

Perceptual grouping enables coherent representations of the external world. Classically, grouping is revealed in Gestalt principles with simple

Correspondence should be addressed to Jelena Ristic, Department of Psychology, McGill University, 1205 Dr. Penfield Ave, Montreal, QC, H3A 1B1, Canada. E-mail: jelena.ristic@mcgill.ca

We would like to thank E. Brock, I. Deschenes, T. Friesen, A. McEvoy, A. Rozenzweig, and N. Zepelowitz for their help with this project, and three anonymous reviewers for their detailed comments on previous versions of this manuscript.

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), the Social Sciences and Humanities Research Council (SSHRC), Stairs Foundation, and William Dawson Funds to J.R., and fellowships from NSERC and SSHRC to J.B.

geometric shapes (Kanizsa, 1979). However, the question of how perceptions are grouped within dynamic contexts remains relatively unexplored. To measure perceptual grouping during complex tasks, Newton and Engquist (1976; see also Kurby & Zacks, 2008) developed a unit marking procedure, also known as an event segmentation task. In this procedure, participants are presented with a short clip depicting an ordinary task, like making a bed, and are asked to press a button when, in their opinion, one meaningful event ends, and another one begins. Despite minimal instructions, participants typically produce remarkably consistent responses and reliably identify specific points in the clip, called the breakpoints, as the boundaries of the underlying perceptual units. Perceptual units indexed by this procedure have been found to form unbreakable wholes (Newton, 1973) and to be organized in a hierarchical manner, with smaller units nesting within the larger ones (Zacks, Tversky, et al., 2001). The formation of unit boundaries is thought to be guided by environmental information that influences and attracts attention, like low-level visual features of motion and colour (Zacks, 2004; Zacks, Kumar, Abrams, & Mehta, 2009; Zacks, Speer, Swallow, Braver, & Reynolds, 2007) as well as conceptual information like goals and causal relationships (Magliano, Miller, & Zwaan, 2001; Speer, Reynolds, & Zacks, 2007). Neuroimaging evidence suggests that some perceptual units are formed automatically. This is typically demonstrated by observing similar patterns of brain activity during initial passive viewing and later overt segmentation task (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004; Zacks, Braver, et al., 2001). Specifically, similarity between the implicit neurophysiological markers (e.g., brain's metabolic or electrocortical activity) of unit boundaries identified during passive viewing with those identified during later behavioural indices of segmentation (i.e., during breakpoints) is taken as an indication of the underlying perceptual parsing automaticity, because neurophysiological markers occur similarly across passive and overt segmentation conditions independently from any task demands (Kurby & Zacks, 2008; Zacks, Braver, et al., 2001).

Based on this past research, we hypothesized that humans interpret social information effortlessly because it is automatically parsed into meaningful units of social perception. To test this idea, we asked participants to perform an event segmentation task while recording their key presses and eye movements, which provided indices of both the overt segmentation behaviour and participants' attentional allocation. Participants were at first asked to passively watch a clip depicting a typical daily situation and were later asked to segment the same clip into social and nonsocial events. If environmental information is organized into units of social perception, we anticipated the following results. First, we expected to find consistency in overt segmentation behaviour across participants, with temporally distinct patterns of social and nonsocial breakpoints. Second, given a wealth of studies indicating the importance of eyes and faces in social communication (Pickett, Gardner, & Knowles, 2004; Smilek, Birmingham, Cameron, Bischof, & Kingstone, 2006), we also expected that attending to eyes and faces would influence the parsing of social but not nonsocial events. Finally, if social event segmentation behaviour was guided by an automatic organization of environmental content into "social" units based on attention paid to available social cues, we also expected to observe a consistent relationship between the indices of attentional allocation (i.e., proportions of fixations) recorded during specifically identified social breakpoints and those same indices recorded at corresponding time points during the initial passive viewing, which approximates natural behaviour (e.g., see Kurby & Zacks, 2008 for similar approaches).

EXPERIMENTAL STUDY

Method

Participants, apparatus, and stimuli

Thirteen undergraduates watched an 8-minute clip, 481 × 681 pixels in size, which was presented at 60 frames per second on a 16-inch monitor at an approximate viewing distance of 57 cm. The clip

was filmed in the laboratory without editing cuts. The lighting was constant throughout the clip. The clip content was as follows (see Figure 1A). In the first 90 s, water boils in the kettle, a call is missed on the phone, and the printer prints a page. Then, at 98 s, Actor 1 enters and reviews his weekly class schedule and needed groceries aloud. Between 157 and 189 s, Actor 1 returns a missed call to his mother and leaves a message. At 198 s, Actor 1 leaves the room but misses another phone call from 211–229 s. Between 240 and 248 s, the printer receives and prints a fax. Actor 1 returns at 253 s and speaks with his mother on the phone about needed groceries. After that, at 320 s, he begins playing a game on his phone. Then, at 338 s, Actor 2 enters, and the two actors begin playing the online Sudoku individually. At 390 s, Actor 2 starts playing a different game (Solitaire), but at 420 s asks to play the same Sudoku game with Actor 1. The two actors then start playing the same Sudoku game online together on the same phone from 420–455 s, talking about the moves. When the game is finished, at 455 s, the two actors continue playing different games individually, and at 475 s Actor 1 falls asleep. Thus, the clip content manipulated whether none, one, or two actors were present and if they were interacting.

Design and procedure

Experiment design closely mirrored the typical event segmentation methodology (e.g., Kurby & Zacks, 2008; Newton & Engquist, 1976; Zacks & Swallow, 2007). All participants viewed the clip three times. At first, in the passive condition, they were instructed to “*Watch the movie carefully and remember as much they can*”. Then, they were asked to segment the clip into nonsocial (“*Press the spacebar when one nonsocial event ends and a new one begins*”) and social events (“*Press the spacebar when one social event ends and a new one begins*”), in a counterbalanced order. This manipulation allowed us to examine whether segmenting based on social information was possible and if so how this process was influenced by the perceptual information contained within the clip. Further, this set-up also allowed us to examine the relative

automaticity of social event segmentation by assessing the similarities between the eye movements observed during passive viewing, which approximate naturalistic behaviour, and those measured during specific breakpoints, which index the perceptual event formation overtly. Increased similarity with passive viewing would indicate more similarity with natural behaviour and suggest underlying parsing automaticity (e.g., Zacks, Braver, et al., 2001; Zacks, Swallow, Vettel, & McAvoy, 2006). Eye movements were recorded using a remote EyeLink 1000 eye tracker, sampling with a temporal resolution of 500 Hz and a spatial resolution of 0.46°. Nine-point calibration and drift correction procedures were performed before each condition. Fixations were detected using a standard SR Research online algorithm with a velocity threshold of 30°/s and acceleration threshold of 8000°/s².

Results

Due to the length and the complexity of the clip, we grouped participants' key presses in 15-s intervals, by summing the number of key presses for all participants in each 15-s bin. Subsequently, we calculated the mean and standard deviation for those values. To determine group-based segmentation behaviour, we defined breakpoint windows as each 15-s interval in which the total number of key presses exceeded 1 standard deviation of the group average for all 15-s windows (e.g., Newton & Engquist, 1976; Zacks et al., 2006). Such group-based event boundaries have been shown to provide stable estimates of overt segmentation behaviour (Zacks et al., 2006). Valid eye movements, or those directed to the clip content (94.5%), were analysed as a proportion of total fixations that fell within one of six mutually exclusive dynamic regions of interest (dROIs; eyes, face, body, hands, background, objects—including phone, printer, and kettle). ROIs were hand drawn for each clip frame while first fixation was coded for each frame and for each breakpoint. The mean proportion of fixations for each dROI was computed by dividing the number of fixations recorded for each dROI by the total number of fixations recorded at that breakpoint. Fixation proportions

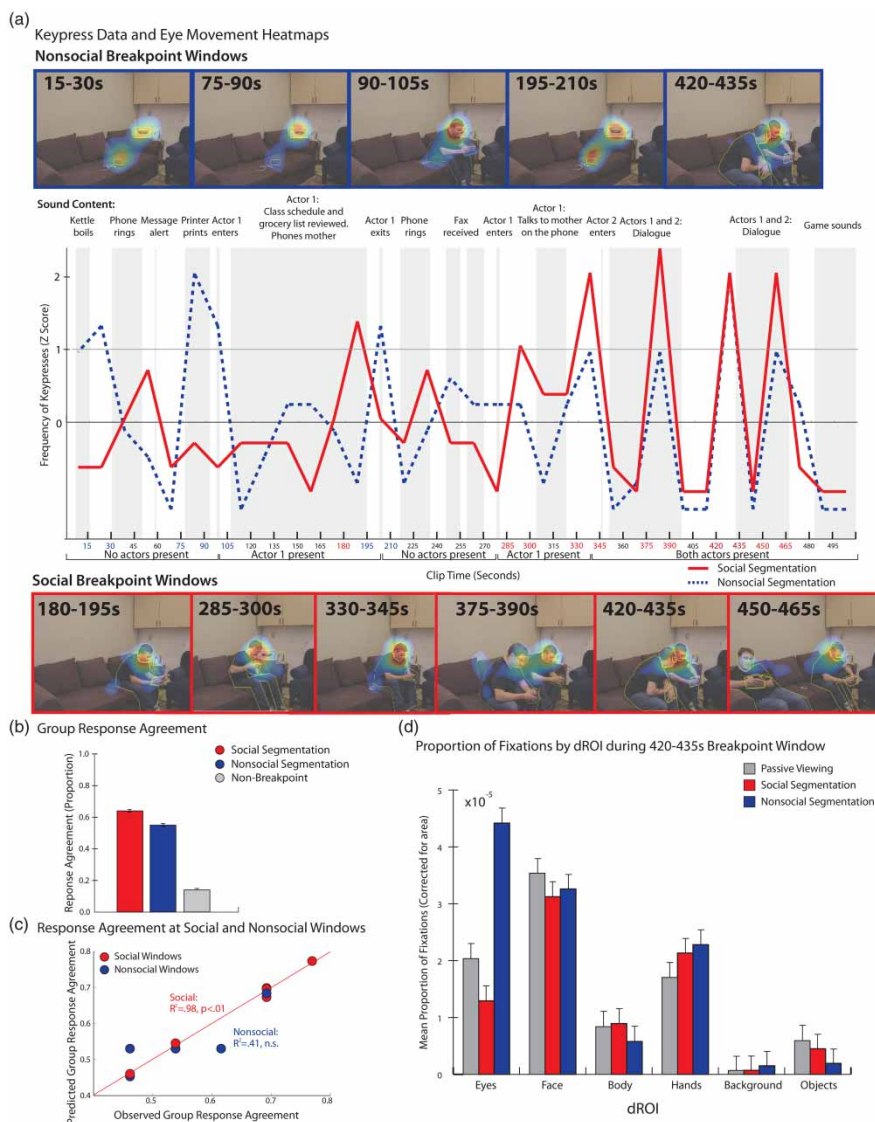


Figure 1. Results. (A) Standardized frequency of key press responses as a function of clip time and segmentation condition as well as eye movement heatmaps showing dynamic regions of interest (dROIs) and fixation density for frames associated with each breakpoint window. Greyed-out areas indicate the parts of the clip containing sound and dialogue. Note that the presence of sound did not influence parsing of social unit boundaries. A reliable relationship between the proportion of time that sound was present within each breakpoint and the participant agreement in indicating that unit boundary emerged for nonsocial, $R^2 = .77$, $F(1, 3) = 14.5$, $p < .05$, but not social breakpoints, $R^2 = .22$, $F(1, 4) = 2.4$, $p > .2$, even though social breakpoints on average contained more sound [social: $t(5) = 4.6$, $p < .05$; nonsocial: $t(4) = 1.9$, $p > .1$, one sample, two-tailed]. (B) Proportion of participants responding at social and nonsocial event boundaries and across all other nonbreakpoint 15-s intervals. Error bars denote standard deviations. (C) The relationship between observed (x-axis) and predicted (y-axis) group response agreement for each social and nonsocial breakpoint, as determined by the parameters from the multiple linear regression models. Reliable regression model includes an estimated linear fit. (D) Area-normalized proportion of fixations falling within each dROI during the overlapping 420–435-s breakpoint window. Error bars represent the standard error between the difference of the means. To view this figure in colour, please visit the online version of this Journal.

were normalized for dROI area by dividing the proportion of fixations for each dROI by the area of that dROI measured in pixels (Birmingham, Bischof, & Kingstone, 2008; Smilek et al., 2006). This procedure is necessary when fixation data are compared across ROIs of different physical sizes (e.g., background vs. eyes), as it adjusts for the fact that larger ROIs might receive more fixations by chance because of their size (see Birmingham, Bischof, & Kingstone, 2009, for a similar analysis). It is important to note though that areas of individual dROIs remained constant within statistical comparisons, as our analyses compared fixations across the same temporal points in the clip—that is, breakpoint windows.

Unique social and nonsocial breakpoints

Our first hypothesis concerned the question of whether social information was grouped into distinct perceptual units. We examined this by comparing the breakpoint window patterns in social and nonsocial segmentation conditions. Our data confirmed this notion, with Figure 1A showing a clear pattern of temporally distinct social and nonsocial breakpoint windows. The time points corresponding to clip times between 180–195, 285–300, 330–345, 375–390, and 450–465 s were identified as uniquely social breakpoint windows, while those corresponding to clip times between 15–30, 75–90, 90–105, and 195–210 s were identified as uniquely nonsocial breakpoint windows. The 420–435 s window was marked as both a social and a nonsocial breakpoint. As shown in Figure 1A, during the uniquely social breakpoint windows, Actor 1 was leaving a message on the phone, talking on the phone, and playing a game on the phone with Actor 2. During the uniquely nonsocial breakpoints, the phone rang, the printer printed a page in an empty room, and Actor 1 read a grocery list to himself.

To rule out the possibility that these breakpoint windows emerged because a few outlier participants responded at a higher rate during breakpoint windows, we analysed the participants' mean proportion agreement in responding during breakpoint windows. We found that participants were remarkably consistent in responding at both social and

nonsocial breakpoint windows, as illustrated in Figure 1B (social breakpoints = 64.1% agreement: $t(5) = 13.5$, $p < .0001$; nonsocial breakpoints = 55% agreement: $t(4) = 12.3$, $p < .001$; both tests one-group, two-tailed, tested vs. hypothesized null, i.e., 0% agreement; the same result is observed when agreement at social and nonsocial breakpoints is compared against average agreement in responding (14.4%) at all other nonbreakpoint windows using paired two-tailed t -tests, both $ts > 5.2$, $ps < .05$). This shows that the breakpoint windows reflected a consistent group-based agreement in segmentation behaviour. Thus, in addition to segmenting social and nonsocial events differently, participants were also remarkably consistent in marking the boundaries of those distinct social and nonsocial events.

To address the second question of whether attending to social information aided the formation of social but not nonsocial events, we next examined the eye fixation data during the uniquely social and nonsocial breakpoint windows and assessed how those attention indices related to participants' response agreement in segmenting social and nonsocial events. First, we considered the proportions of fixations directed to different dROIs during social and nonsocial breakpoints. Figure 1A illustrates fixation density for each social and nonsocial breakpoint. It shows that during social breakpoints, participants fixated on eyes and faces, while during nonsocial breakpoints they fixated on objects, probably because most nonsocial breakpoints contained no persons, as we discuss later on. The interparticipant mean proportions of fixations that fell in each dROI were examined using two repeated measures analyses of variance (ANOVAs), which were conducted separately for social and nonsocial breakpoints and included the breakpoint windows (social: 180–195, 285–300, 330–345, 375–390, 450–465 s; nonsocial: 15–30, 75–90, 90–105, 195–210 s) and dROI (eyes, face, body, hands, background, objects) as factors. During social breakpoint windows, most fixations were directed at eyes and faces [$F(5, 60) = 6.38$, $MSE = 2.381 \times 10^{-9}$, $p < .0001$] especially during the 285–300 s breakpoint window [dROI \times Breakpoint Window; $F(20, 240) = 3.6$, $MSE = 1.125 \times 10^{-9}$, $p < .0001$],

when Actor 1 speaks with his mother on the phone. In contrast, during nonsocial breakpoints, most fixations were directed at objects [$F(5, 60) = 5.43$, $MSE = 8.827 \times 10^{-5}$, $p < .0001$], especially during the 195–210 s breakpoint window [dROI \times Breakpoint Window; $F(15, 180) = 5.42$, $MSE = 8.819 \times 10^{-5}$, $p < .0001$], when Actor 1 leaves the room. Thus, in general, when actors were present, participants fixated on their eyes and faces, and the time window was marked as a social event boundary. When no actors were present however, participants fixated on objects, and the time window was marked as a nonsocial event boundary.

Second, we examined whether paying attention to social information was related to overt segmentation behaviour. If so, it would be reasonable to hypothesize that the mean proportion of fixations recorded for social dROIs during the breakpoint windows, which index attentional allocation, would be reliably related to group response agreement, which indexes overt segmentation behaviour at those breakpoint windows. Supporting this hypothesis, our analyses revealed that paying attention to social dROIs reliably predicted group response agreement in marking social breakpoint windows but not nonsocial ones. Multiple regressions analyses were run with the mean proportion of fixations recorded for each dROI included as predictor variables of the group response agreement at social and nonsocial breakpoints separately. For social breakpoints, the model that included the proportions of fixations directed to eyes, face, and body dROIs accounted

for 98% of variance in the group response agreement in marking the boundaries of those social events [$R^2 = .98$; $F(3, 2) = 92$, $p < .01$]; however, the same predictors did not account for a reliable amount of variance observed in the group response agreement for marking the boundaries of nonsocial events [$R^2 = .41$; $F(2, 2) = 2.4$, $p > .2$]. In fact, in addition to not being reliably related to the proportions of fixations directed to social dROIs, group response agreement at nonsocial breakpoint windows was not reliably related to proportions of fixations directed to any dROI [all R^2 s $< .46$, p s $> .1$], including those directed to nonsocial dROIs, i.e., Background and Objects [all R^2 s < 1 , F s < 1.5 , p s $> .4$]. Figure 1C shows the relationship between the group response agreement (i.e., the proportion of participants responding) that we observed in our experiment (x -axis) and the group response agreement predicted by the regression models including eyes, face, and body dROIs as predictors (y -axis) for social and nonsocial breakpoints separately. Table 1 shows the parameters of the two multiple regression models.

Thus, our data also suggested that paying attention to social information was related to group agreement in marking social events. That is, a larger amount of variance in participants' response agreement during social segmentation relative to nonsocial segmentation condition was accounted for by variance in the proportion of fixations (i.e., attentional allocation) directed to dROIs containing social content. No significant links between the proportions of fixations and response

Table 1. Multiple regression results and parameters

Model Fit Variables	Social breakpoints						Nonsocial breakpoints					
	$R^2 = .993$; Adjusted $R^2 = .982$						$R^2 = .706$; Adjusted $R^2 = .412$					
	<i>B</i>	<i>SE B</i>	β	<i>t</i>	<i>p</i>	<i>VIF</i> **	<i>B</i>	<i>SE B</i>	β	<i>t</i>	<i>p</i>	<i>VIF</i>
Constant	.876	.043	–	20.5	.002	–	.538	.004	.	12.1	.007	–
Eyes	237	481	.081	<1	ns.	7.5	6230	3304	1.2	1.9	ns.	2.8
Head	73.7	1436	.008	<1	ns.	6.3	–3723	4299	–56	<1	ns.	2.8
Body*	–23357	1835	–1.05	–12.7	.006	1.9	–	–	–	–	–	–

Note: SE B = standard error B; VIF = variance inflation factor.

*Excluded variable. **Combined face and eye dROI ($R^2 = .993$; Adjusted $R^2 = .988$); VIF = 1.8.

agreement emerged for the nonsocial segmentation task. This suggests that participants attended to social information more consistently while performing the social segmentation task, which in turn was related to the high group-based agreement in marking the boundaries of uniquely social events. It is important to note here that although regression analyses were based on a relatively small sample and should be interpreted with caution, both models included the same number of parameters but resulted in divergent results. Thus, it is unlikely that the absence of a reliable relationship in the nonsocial segmentation condition reflects low power in that analysis.

Taken as a whole, the analyses conducted on unique social and nonsocial breakpoints support the conclusion that attending to social information is related to the formation of social perceptual unit boundaries. Furthermore, they also suggest that the presence of people might be an important determinant of whether dynamic information is grouped into social events, possibly denoting the highest level of structural hierarchy of social event segmentation. We test these conclusions further in the following section in which we analyse the breakpoint window that was marked as both a social and nonsocial event boundary.

Overlapping social and nonsocial breakpoint

In addition to revealing the temporally unique social and nonsocial breakpoints, our results also indicated that the 420–435 s breakpoint window, when the two actors were playing a game together, was marked as both a social and a nonsocial event boundary. Analysing this breakpoint window provided us with an opportunity to delineate the types of visual information that are used in parsing social and nonsocial events on a more fine-grained level and additionally facilitated further tests of automaticity of social event segmentation by investigating the links between attention indices observed during the initial passive viewing with those observed during this overlapping breakpoint. Furthermore, the inspection of fixation patterns during this overlapping breakpoint also allowed us to examine the hierarchical structure of social events, as the clip content

during this breakpoint window was now equated in terms of the number of individuals present in the scene.

Based on the large literature indicating attentional priority for eyes and faces (e.g., Frischen, Bayliss, & Tipper, 2007; Smilek et al., 2006), one might predict that parsing social events on a more fine-grained scale, as it would be required during a breakpoint window that is marked as both social and nonsocial, would depend on the information conveyed by subtle social cues. If so, during this overlapping breakpoint window, participants should be fixating eyes and faces more while performing a social than a nonsocial segmentation task. Moreover, if paying attention to social information was linked with segmenting social events, fixation patterns during the initial passive viewing, which approximate naturalistic behaviour, should be associated with fixation patterns from the later social segmentation task. Figure 1D plots the proportions of fixations for each dROI for passive viewing, social segmentation, and nonsocial segmentation conditions. A repeated measures ANOVA examined these data as a function of the viewing condition (passive, social, nonsocial) and dROI (eyes, face, body, hands, background, objects) included as within-subjects variables. Overall, and supporting the existing literature, most fixations were directed to social dROIs of eyes, faces, and hands [$F(5, 60) = 9.5$, $MSE = 6.901 \times 10^{-10}$, $p < .0001$]. This result held similarly for passive viewing and social segmentation conditions. However, during the nonsocial segmentation, participants fixated the eye region most frequently [dROI \times Viewing Condition: $F(10, 120) = 2.02$, $MSE = 3.196 \times 10^{-10}$, $p < .05$]. This suggests that in addition to attending to social cues similarly while passively watching the clip and while segmenting it into social events, participants also attended to social cues while they were parsing nonsocial events. Thus, when the clip content was equated with respect to the number of individuals present in the scene, social events were parsed based on the information indicated by more subtle social cues, conveyed by faces and eyes, rather than just the coarse ones, like presence of people. In addition, while fixation

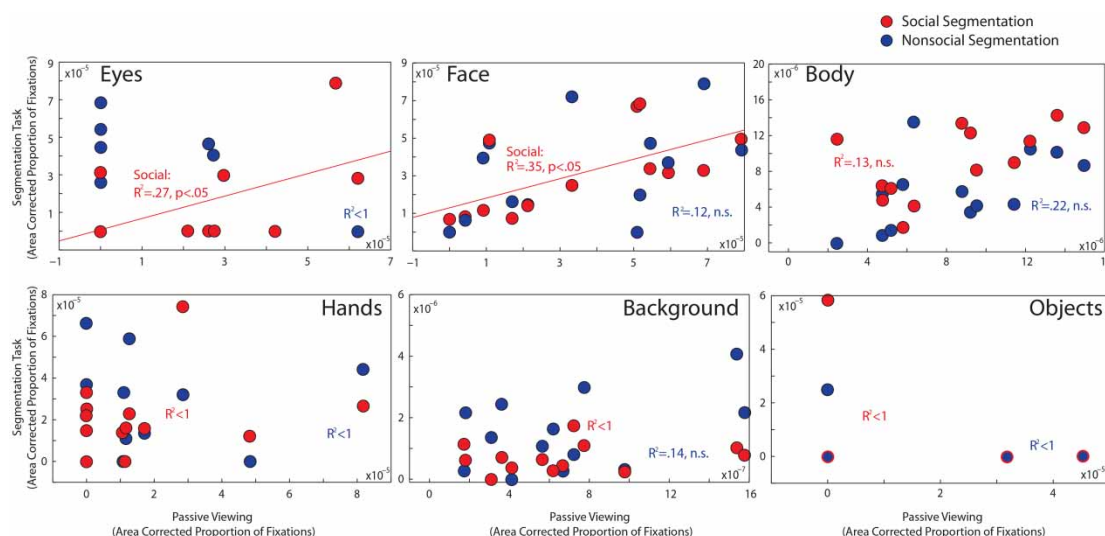


Figure 2. The relationship between the proportions of fixations recorded during passive viewing and during overt segmentation tasks. Scatterplots show the individual participant's area-corrected proportion of fixations recorded during passive viewing (x-axis) and the corresponding proportions of fixations recorded during social and nonsocial overt segmentation tasks (y-axis), as a function of dynamic regions of interest (dROI). The figures and analyses are based on individual participants' nonzero fixations recorded for each dROI. Reliable regression models include an estimated linear fit. To view this figure in colour, please visit the online version of this Journal.

patterns for dROIs mirrored one another across the passive and social conditions, an increased proportion of fixations landed to the eye region during the nonsocial segmentation task. This is a surprising result that may indicate key distinguishing features between the processes that underlie social and nonsocial event segmentation. We return to this point in the Discussion.

Finally, to more directly examine the automaticity of social event segmentation, we investigated the relationship between attentional allocation during passive viewing and attentional allocation during overt social and nonsocial segmentation tasks. As a reminder, this analysis suggests automaticity of the underlying segmentation process by revealing the similarities between the implicit attentional markers of perceptual segmentation during initial passive viewing with those same indices observed during later overt tasks (see Kurby & Zacks, 2008; Zacks, Braver, et al., 2001, for similar analyses). As such, it allows one to confirm that the environmental features that have been found important for overtly identifying

social breakpoints are the same as those that are looked at naturally during passive viewing.

Thus, if this dynamic clip is automatically parsed into units of social perception based on attention paid to social cues, the proportions of fixations falling within the individual dROIs during passive viewing should be reliably related to the proportions of fixations falling within those same dROIs and specific time points during the social segmentation task but not during the nonsocial segmentation task. Put simply, attention indices during the initial passive viewing should be similar to attention indices during the later social segmentation task and not similar to attention indices during the nonsocial segmentation task.

We tested this hypothesis using linear regression analyses. Here, individual participants' mean proportion of fixations for each dROI during passive viewing was entered as a predictor of their individual proportion of fixations recorded in those same dROIs and temporal windows during social and nonsocial segmentation tasks. Separate regressions were run for each dROI and segmentation (social,

nonsocial) task condition. The relationships between the individual participants' proportions of fixations during passive viewing and during the two segmentation tasks are shown in Figure 2 for each dROI. Confirming our hypothesis, the proportions of fixations that were directed to eyes and face dROIs during the initial passive viewing [eyes dROI: $R^2 = .27$, $F(1, 11) = 5.387$, $p < .05$; face dROI: $R^2 = .35$, $F(1, 11) = 7.396$, $p < .05$; all other $R^2 < .2$, $ps > .1$] reliably predicted the proportions of fixations directed to those same dROIs during social segmentation. In sharp contrast, the same relationship did not hold for the nonsocial segmentation condition. Here, the proportions of fixations directed to eyes and face dROIs during passive viewing did not predict the proportions of fixations observed in those same dROIs during nonsocial segmentation. In fact, no reliable relationship was found between the proportion of fixations recorded during passive viewing and nonsocial segmentation for any dROI [face dROI: $R^2 = .12$, $F(1, 11) = 2.631$, $p > .1$; eyes dROI: $R^2 = .01$, $F < 1$; all other dROIs: $R^2 < .22$, $ps > .6$].

Dovetailing with our analyses of unique breakpoint windows and conceptually replicating past studies that have investigated attentional effects elicited by static social information (e.g., Friesen & Kingstone, 1998; Smilek et al., 2006), these data indicate that dynamic social information also appears to be attended automatically. Moreover, our results also suggest that attention paid to social information consequently aids the formation of social perception units, as eye movements directed to social cues during passive viewing reliably predicted eye movements towards the same social cues during the social segmentation task. In contrast, no reliable relationship emerged between the eye movements directed to any dROI observed during the initial passive viewing and those recorded the later nonsocial segmentation task, suggesting a possible involvement of more strategic processes during nonsocial segmentation.

Control analyses

It is important to note that our data do not reflect artefacts stemming from the eye tracking

measurements or data handling procedures. To verify this, we conducted four different control analyses. The first one examined whether the actual eye tracking spatial resolution accuracy allowed for unambiguous analyses of all dROI data, and especially for fixations directed to the smallest eye region. An examination of all participants' calibration data indicated that the average tracking spatial resolution of $0.463^\circ \pm 1.8^\circ$ was sufficient to resolve the size of the eyes dROI, which averaged between 0.5° and 1° of visual angle. Thus, the eye tracking parameters allowed for the measurement of all fixation data. Nevertheless, at a suggestion from a reviewer, we also analysed our data by combining the fixation data for eyes and face dROIs. All of the results that we report here held except that the interaction between the viewing condition (passive, social, nonsocial) and the combined face/eye dROI during the overlapping breakpoint was now reliable only marginally, $F(8, 96) = 1.8$, $p = .07$. This probably reflects averaging between a large difference in the proportion of fixations observed between the original eye and face regions in the nonsocial segmentation task and a small difference between the proportions of fixations between those two same dROIs in the social segmentation task (see Figure 1D). However, even in this conservative test, a strong trend towards an increased proportion of fixations to the combined face/eyes dROI during nonsocial segmentation remained.

The second control analysis examined whether the increase in the proportions of fixations directed to the eye region in the nonsocial segmentation task during the overlapping breakpoint window reflected the differential frequencies of fixations that may have occurred during social and nonsocial tasks. One alternative possibility is that this surprising result may simply reflect more fixations in the nonsocial condition. To examine this, we inspected the number of raw valid fixations in each social and nonsocial segmentation tasks during the overlapping breakpoint. We found that the total number of raw fixations for social and nonsocial segmentation conditions did not differ (271 vs. 297 in the social vs. nonsocial condition, respectively), $t(12) = -1.1$, $p > .3$, paired t -test, two-tailed),

strongly suggesting that any differences between the two conditions in the proportions of fixations did not reflect differential frequencies in the number of fixations.

In a related third analysis, we examined whether raw fixations were distributed evenly across all clip frames comprising the overlapping breakpoint in both segmentation tasks. We did so because one might argue that the differences in the proportions of fixations directed to the eye region during social and nonsocial segmentation tasks reflected the differential temporal distribution of fixations for those two conditions during the overlapping breakpoint window. Our analyses did not support this alternative. When we visually inspected the distribution of raw valid fixations within this overlapping breakpoint window, we found that fixations for social and nonsocial tasks were distributed evenly across all clip frames, rather than showing temporally distinct clusters. Furthermore, the mean numbers of raw fixations that were directed to the most pertinent dROIs of eyes and faces were also distributed evenly across clip frames and were moreover equated in frequency across social and nonsocial segmentation tasks. That is, we found no differences in the mean number of raw fixations that occurred in each of the five bins that contained the fixation counts for 59 frames (17 vs. 19 fixations per bin), $t(4) < 1$, $p > .5$, two-tailed, paired.

Our final control analyses confirmed that the increase in the proportions of fixations directed to the eye region in the nonsocial segmentation condition during the overlapping breakpoint was not an artefact of the area-normalization procedure. First, the size of the eye dROI was identical across the two segmentation conditions, and thus the difference in fixation proportions cannot reflect the physical ROI size changes. Second, as noted above, there was no general increase in the number of fixations between the two segmentation conditions, which yielded nearly identical numbers of total correct fixations. Finally, this result held in the raw fixation data as well, as the only reliable difference across individual dROIs in raw fixation counts between social and nonsocial conditions emerged for eyes dROI, which received reliably more fixations in the nonsocial condition,

$t(12) = -2.66$, $p < .05$, two-tailed (all other t s < -1.8 , p s $> .08$).

Discussion

By using a modified unit marking procedure here we demonstrate for the first time that humans naturally organize dynamic information into units of social perception. Three lines of evidence support this conclusion. First, key press data indicated clear patterns of unique social and nonsocial breakpoints and a high between-participant agreement in marking those units. This finding conceptually replicates the existing data (Newton & Engquist, 1976; Zacks et al., 2007) and further extends event segmentation methodology into a social domain. Second, eye movement data further showed that identifying the boundaries of social events was associated with an increased attentional allocation to actors' eyes and faces, demonstrating the importance of eyes and faces when processing social content. Finally, linking these data with natural behaviour, the comparisons between passive viewing and the segmentation tasks during the overlapping breakpoint window indicated that fixations falling to social dROIs (i.e., eyes, face) during passive viewing reliably predicted fixations falling to those same dROIs during social but not nonsocial segmentation tasks. These results conceptually replicate and extend past behavioural (e.g., Newton & Engquist, 1976; Zacks & Swallow, 2007) and neuroimaging data (e.g., Kurby & Zacks, 2008; Zacks, Braver, et al., 2001), indicate automaticity of attending to social cues during dynamic situations, and suggest hierarchical organization of social events. Taken together, our data reveal a perceptual organization process that may be one of the driving forces behind expert human social behaviour and additionally provide a novel methodology for linking typical and atypical social behaviour with the underlying perceptual and cognitive mechanisms. We discuss each of these points next.

Although using a complex dynamic clip portraying social and nonsocial content, our results extend both the data from the past social attention studies and those from event segmentation investigations.

Like studies on social attention, which utilized both simple schematic cues (e.g., Friesen & Kingstone, 1998) and more complex naturalistic stimuli (e.g., Birmingham et al., 2008), we found that participants preferentially attended to social information conveyed by eyes and faces in the absence of any explicit instructions to do so. Furthermore, we also found that attention paid to social information was reliably linked with participants' agreement in segmenting social events, which along with the high agreement in responses during social breakpoints suggests that social event segmentation is guided by the attentional allocation toward social cues in the environment. Our data also suggest that in contrast to nonsocial events, social events appear to be parsed automatically based on attentional allocation to social content. Using a similar logic to that in past research, we demonstrated this by showing that attentional selection during initial passive viewing reliably predicted attentional selection during later overt social segmentation but not nonsocial segmentation behaviour.

It has been suggested to us that this result might not be indicative of automatic processes in parsing of social information because overt segmentation behaviour was elicited using instructions to parse the clip content. We believe that this is not true for three reasons. One, participants were never explicitly instructed to pay attention to any particular part of the clip or a specific visual feature within a clip during the segmentation tasks. They were simply instructed to "press a key when one (social or nonsocial) event ended and another one began". Thus, participants' overt segmentation behaviour was guided by the information available in the environment (i.e., the clip), rather than the task set created by the instructions. Two, if no systematic underlying processes guided participants' overt segmentation behaviour, one might expect that each observer would utilize different environmental information, resulting in large key press variability across participants and consequently no reliable between-participant agreement in marking the unit boundaries. This was not the case in our data, as we found not only that participants were consistent in marking the boundaries of social and nonsocial units, but that they also did so with

a remarkable between-participant agreement in responding. Furthermore, between-participant agreement during social event segmentation was also reliably linked with paying attention to social information. Three, if social event segmentation was not related to an underlying automatic allocation of attention to social cues, no similarities in fixation patterns between passive viewing, which approximates the naturalistic viewing conditions, and overt social segmentation task should emerge. This was not the case in our data either, as we showed systematic links between attention during passive viewing and attention during social segmentation. Note that since these analyses compared the implicit indices of attentional behaviour during passive viewing (i.e., proportion of fixations), recorded *before* any overt tasks, with those same indices recorded during the *later* segmentation behaviour, one can be confident that the data from passive viewing are never contaminated by any task set. Existing neuroimaging evidence supports this notion and shows that specific and transient metabolic and electrocortical changes in the extrastriate visual perception areas are time-locked to the event boundaries similarly during passive viewing and overt segmentation conditions (Kurby & Zacks, 2008). Likewise, we found that the proportions of fixations that fell to eyes and face regions during passive viewing reliably predicted the proportions of fixations that fell to those same dROIs and time points within the clip during social but not nonsocial breakpoints. Thus, our data demonstrate that the implicit indices of participants' attention (i.e., their fixations) exhibit a great deal of similarity during naturalistic passive viewing and during overt social segmentation. Importantly, the same similarity between the attentional processes during passive viewing and during overt nonsocial segmentation was not found, once again supporting the notion that *social* event segmentation involves similar automatic processes as naturalistic behaviour recorded in the absence of any task.

In contrast to social segmentation, however, it appears that the nonsocial event segmentation task involved more controlled processes, which were employed explicitly in response to task

demands in order to evaluate the available perceptual information. Because explicit behaviour is dependent on task set, it is reasonable to expect some similarity in behaviour across participants in the nonsocial segmentation task. This was supported by our key press data, which indicated reliable participant agreement in overt nonsocial segmentation behaviour. However, a task-dependent process should not approximate naturalistic, spontaneous behaviour recorded during passive viewing because this initial behaviour did not depend on the later task instructions. This was supported by our fixation data. Fixation patterns during the nonsocial segmentation task were not reliably related to the proportion of fixations that fell to either social or nonsocial regions during passive viewing, indicating that overt parsing of nonsocial information occurred less automatically. Curiously, however, our data also indicated that an increased proportion of fixations fell within the eyes region during the nonsocial segmentation task. This shows that while segmenting social events might depend on automatic selection of social cues, segmenting nonsocial events might require extracting extra information from similar perceptual sources (e.g., eyes), which in this case might be employed explicitly rather than automatically in order either to eliminate the possibility of social content or to garner more information needed to make a response decision. In other words, in addition to providing a wealth of social knowledge, eyes also appear to communicate critical information that participants may use explicitly to determine whether an event, which is embedded in an otherwise social scene, might be nonsocial. As such, the observed increase in the proportion of fixations to the eye region during nonsocial segmentation most likely reflects controlled processes that are at play during complex situations in which social and nonsocial events overlap.

Mirroring the typical event segmentation results, our data also revealed a potential hierarchy in the structure of social events. Specifically, we found that the boundaries of uniquely social events were related to the presence of people, but that the boundaries of more complex events containing both social and nonsocial information

were guided by the information conveyed by specific social cues (i.e., those indicated by eyes and faces). That is, during the more complex overlapping breakpoint, attention paid to subtle social cues indicated by faces and facial features was the key determinant of social segmentation. As such, this simple yet powerful finding suggests that at the highest level of hierarchy, the units of social perception may be formed based on the attentional priority for detecting the presence of people, but that on a more fine-grained level, they are influenced by the attentional priority assigned to the information conveyed by social cues indicated by people's eyes and faces. This dovetails with the existing data indicating that event segmentation is guided by both general conceptual information as well as specific cues in the environment (e.g., Magliano et al., 2001; Zacks et al., 2006).

Taken as a whole then, our data open several possible future research venues. One, investigations into the neural mechanisms that underlie social event segmentation might indicate the extent to which selection of social information guides social and nonsocial event segmentation automatically. Such investigations might also offer an insight into the nature of social information that is utilized while parsing social events. Two, studies aimed at delineating the precise temporal structure of social events, and particularly those that are marked as both social and nonsocial, would offer a more detailed analysis of factors that may determine social segmentation. In contrast to previous studies, in which participants were typically asked to parse 1-min-long clips depicting simple procedural behaviours (e.g., making a bed), here we examined how participants parsed a longer 8-min video clip, which depicted both complex conceptual content and simple procedural content. Due to the brevity of the clip content, past studies provided a precise temporal understanding of segmentation behaviour by examining key press responses at each 1-second interval. In contrast, here we examined segmentation behaviour during 15-second temporal windows, which resulted in a less precise temporal understanding of social event parsing. However, the present analysis of breakpoint windows was based on a similar number of

temporal intervals as used by previous research, and as such afforded similar statistical power to detect event boundaries. Analysing responses on a 15-s time scale also provided a logical temporal breakdown of the entire clip according to the typical duration of events that unfolded. By broadening the breakpoint window, we were thus able to examine how more complex social and nonsocial content may be grouped during extended periods of time. That is, we were able both to capture the coarse social information that is used in identifying social events (i.e., presence of people) as well as to understand the influence of the more specific social cues (i.e., fixations directed to faces and eyes) in identifying social events during complex scenes that contained both social and nonsocial content. In addition to revealing novel results about social event segmentation, our results replicated prior segmentation data in terms of the hierarchical organizations of events and past social attention data in terms of the prioritization of social cues. Future investigations would benefit from analyses conducted on a more fine-grained time scale—that is, based on individual key presses occurring within each social and nonsocial breakpoint window and/or across whole shorter clips to reveal a more precise temporal relationship between the environmental information used to parse social and nonsocial events. Those analyses would also have the potential to uncover the speed of the underlying processes that contribute to social event segmentation and social event content updating as well as to provide more insight into the differences between social and nonsocial event segmentation. Furthermore, analyses of the content of social and nonsocial perceptual units would offer additional knowledge into the cognitive and perceptual processes that guide unit perception versus those implicated in unit segmentation.

Finally, one might utilize social event segmentation methodology to understand atypical social behaviour and to study the range of social competence in typical populations. One might predict that individuals with attentional difficulties (i.e., attention-deficit/hyperactivity disorder, ADHD) or persons with social dysfunctions (i.e., autism) may employ different underlying processes to parse the

environmental content (e.g., Ristic et al., 2005), ultimately resulting in differential grouping of social events, misinterpretations of attentional cues, social situations, and inappropriate behaviour enactments (e.g., Courtney & Cohen, 1996). Furthermore, irregularities in social perception might also be observed within typically developing individuals and as such provide a link between social perception and a range of typical social and attentive behaviours.

In conclusion, our study provides evidence that social information in the environment guides parsing of complex dynamic situations in an automatic manner. This methodology offers a novel way for experimentally linking basic perceptual and attentional processes at group and individual levels and a new way to measure and characterize social attention and the resulting more complex social processes in both typical and clinical populations.

Original manuscript received 23 June 2014

Accepted revision received 4 September 2014

First published online 7 November 2014

REFERENCES

- Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Social attention and real-world scenes: The roles of action, competition and social content. *The Quarterly Journal of Experimental Psychology*, 61(7), 986–998.
- Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Get real! Resolving the debate about equivalent social stimuli. *Visual Cognition*, 17(6), 904–924.
- Courtney, M. L., & Cohen, R. (1996). Behavior segmentation by boys as a function of aggressiveness and prior information. *Child Development*, 67(3), 1034–1047.
- Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, 24(6), 581–604.
- Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, 5(3), 490–495.
- Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, 133(4), 694–724.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical

- activity during natural vision. *Science*, 303(5664), 1634–1640.
- Kanizsa, G. (1979). *Organization in vision: Essays on gestalt perception*. New York: Praeger.
- Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, 387, 767–768.
- Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2), 72–79.
- Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, 15(5), 533–545.
- Newton, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28(1), 28–38.
- Newton, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, 12(5), 436–450.
- Pickett, C. L., Gardner, W. L., & Knowles, M. (2004). Getting a cue: The need to belong and enhanced sensitivity to social cues. *Personality and Social Psychology Bulletin*, 30(9), 1095–1107.
- Ristic, J., Mottron, L., Friesen, C. K., Iarocci, G., Burack, J. A., & Kingstone, A. (2005). Eyes are special but not for everyone: The case of autism. *Cognitive Brain Research*, 24(3), 715–718.
- Smilek, D., Birmingham, E., Cameron, D., Bischof, W., & Kingstone, A. (2006). Cognitive ethology and exploring attention in real-world scenes. *Brain Research*, 1080(1), 101–119.
- Speer, N. K., Reynolds, J. R., & Zacks, J. M. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science*, 18(5), 449–455.
- Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, 28(6), 979–1008.
- Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., ... Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, 4(6), 651–655.
- Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition*, 112(2), 201–216.
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, 133(2), 273–293.
- Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, 16(2), 80–84.
- 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, 150–162.
- Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, 130(1), 29–58.