Sources of Information in Human Action

Jeff Loucks (jloucks@uoregon.edu)
Department of Psychology, 1227 University of Oregon, Eugene, OR, 97403 USA

Dare Baldwin (baldwin@uoregon.edu)
Department of Psychology, 1227 University of Oregon, Eugene, OR, 97403 USA

Abstract

Despite the importance of action identification and discrimination in social cognition, little research has explored how these tasks are accomplished. We investigated whether adults capitalize on local and global sources of information when discriminating human actions. The results of two experiments indicate adults are sensitive to both sources of information in action but selectively attend to local over global information. The findings also parallel what is known about face processing: processing of global information in action is disrupted by inversion, whereas processing of local information is especially reliant on high spatial frequencies.

Keywords: Human action; intentions; selective attention; face processing; discrimination; information processing.

Every day we are witness to the exceptionally complex stimulus that is other people's behavior. In the face of such complexity, it is striking how readily we make sense of others' actions, not only registering the surface form of such action, but also gleaning underlying intent. Processing action in this way is essential for everyday functioning and social interaction.

Given the centrality of action processing to everyday cognitive and social functioning, it is surprising that there has been little research on the fundamental task of identifying and discriminating human actions. Discrimination is a basic component of action categorization. Although every instance of human action is a unique event, it is likely actions are treated as category members in our processing of the behavior stream. Rapidly discriminating actions on the category level is a skill employed countless times per day in everyday life. Moreover, imitation – a hallmark of human social cognition – requires a finely-tuned analysis of the behavior stream to discriminate between different functional actions.

Research on people's discrimination of point-light action displays has demonstrated that low-level visual information can be used to discriminate some basic actions (Johansson, 1973). However, we still know little about people's everyday action processing skills. Processing everyday action may differ from processing point-light action, as such events are more complex, unfold over longer periods, and contain additional information about objects and settings. Given the rich information present in everyday action, it would be beneficial to assess what information people are sensitive to in the context of such action.

Our goal with the current research was to explore basic kinds of information people extract from the motion stream for discriminating between everyday, dynamic human actions. What kind of information do people key on, for example, in discriminating a slap from a caress? Although one could analyze the physical motion differences between these two actions, the number of possible points of comparison is immense. Given this computational complexity, it is not at all self-evident which among these physical differences observers actually track and register in the context of actual action processing. With little previous research to guide us, we looked to another domain of research that has explored similar questions regarding sources of information utilized in identification and discrimination – the domain of face processing.

Faces and Action

Research suggests people are sensitive to two main sources of information in faces: salient, local information, and more global, relational information (Maurer, Le Grand, & Mondloch, 2002). Local information is characterized as information about highly defined, local elements of the face, such as the eyes, nose and mouth (sometimes called "features"). In contrast, global information is characterized as information about the spatial relations among these local elements.

Faces and action have much in common: processing of both seems to require minimal effort on our part, and efficient processing of both has important consequences for social cognition. Given these similarities, we questioned whether faces and action might also share similar processing characteristics. In particular, we were interested in whether people use local and global sources of information to process human action similar to those used to process faces. Some have argued that this dual processing approach may not be specific to faces, and may develop whenever stimuli are processed with expertise (Diamond & Carey, 1986; Gauthier & Tarr, 1997).

In the face processing literature, there is evidence that local and global sources of information are relatively distinct from one another in that they are affected in predictably different ways by different visual manipulations. One such line of evidence is the "face-inversion effect" – the finding that faces are difficult to recognize when turned upside down (Yin, 1969). Research indicates that inversion specifically disrupts the processing of global information while having little to no impact on the processing of local information (Freire, Lee, & Symons, 2000). Presumably the standard global face template people make use of during upright processing is less accessible in the inverted stimulus, and thus recognition and discrimination suffer. Similarly, spatial frequency can also distinguish the processing of these two sources of information. High spatial frequencies encode visual information about local details in an image, on which identification of local facial elements is thought to depend. Thus, processing of local information suffers at low spatial frequencies while processing of global information is unaffected (Goffaux et al., 2005).

A wealth of research indicates that local and global aspects of the face are distinct sources of information for identifying and discriminating faces. If processing action relies on mechanisms similar or related to those involved in face processing, then local and global information in action might be meaningfully distinct from one another, as well as similarly sensitive to orientation and spatial frequencies. In order to explore these issues, we first must define the form local and global information might take in action.

Local and Global Information in Action

Although action seems inherently relational in nature, certain aspects of the motion stream might nonetheless be processed in a relatively local manner. People's judgments about the identity of an action are often based on highly relevant characteristics which are quite local in nature. For example, identifying an action as a grasp relies on identifying a particular kind of hand shape: open enough to surround an object yet closed enough to gain an adequate grip. A slightly different hand shape would result in a judgment that this was a touch or a push instead. When local elements are key to identity judgments across many processing contexts, it seems plausible to consider this an important source of information in processing action. This characterization is consonant with the characterization of local information in the face, as local facial elements are detailed portions of the face that display high functional relevance across many contexts.

Other aspects of the motion stream that are more global and relational in nature seem less key to judging the identity of action, but at the same time not entirely irrelevant to such judgments. In the case of a grasping action, for example, sizable global changes could be made in the trajectory of the reach without affecting the judgment that this was a grasping action. However, the trajectory of motion is a crucial component of the grasp, and alterations in this global property do inform goal identification. For example, processing the trajectory of a grasp is key to identifying the specific goal of a grasp. These global aspects of motion might plausibly be termed relational in that they represent how local action elements are spatially related to one another, much as global information has been conceptualized in the face.

For purposes of this research, then, we defined local information in action as local detail central to identification of small-scale actions. Small-scale actions, such as grasp, place, push, and pull, are the smallest units of action that observers readily identify when asked to segment human action (Newtson, 1973). Take, for example, the actions involved in moving a book. If asked to segment this scenario into the smallest meaningful units, observers would typically mention grasping the book, lifting it, and placing it down in a new location. Identifying the small-scale action unit of grasping requires local detail about a particular hand shape in relation to the object. Thus, we conceptualized local information as highly relevant details that observers might rely on to categorize instances of small-scale action units and differentiate them from other small-action units. Global information, in contrast, was conceptualized as global spatial properties of the motion stream manifested across multiple local elements, such as the overarching trajectory of motion. Pursuing the book-moving example, global information would be information about what path the hand took through space on its way to the book.

Conceptualized this way, one can imagine that processing global action information might rely heavily on a standard global template for human action (head above the arms, legs below the waist, etc.), similar to the way in which a global face template has been implicated in face processing. If so, inverting action might radically undercut the ability to track global information in action, as it has been shown to do in face processing. Similarly, processing of local action information might be especially reliant on high spatial frequencies relative to global action information.

To investigate the meaningfulness of the local versus global distinction in action discrimination, we developed action analogs to the discrimination methodologies utilized in face-processing research (e.g., Freire et al., 2000; Goffaux et al., 2005). Across both experiments, adults watched pairs of videos of everyday actions and made same/different judgments about these pairs. On different trials, videos with alterations in either local action information or global action information were paired against the standard action video. Local change videos were identical to the standard video except for a single alteration in the local detail of a small-scale action. Global change videos were also identical to the standard video except for a single change in the global spatial properties of the motion. In Experiment 1 we investigated whether processing of global information is especially reliant on viewing motion in an upright orientation. In Experiment 2 we explored whether local information is especially reliant on high spatial frequencies.

Experiment 1

In Experiment 1 we investigated whether inversion might selectively impair processing of global information in action. We predicted that observers would be sensitive to both local and global changes when action was presented in its normal, upright orientation but that inversion would disrupt accuracy for pairs differing solely in global information to a significantly greater extent than pairs differing solely in local information.

Method

Participants Thirty-nine adults – all University students (17 male, 22 female) – received partial course credit for their participation.

Table 1: The eight video scenarios used in Experiments 1 and 2.

Scenario	Standard Video	Local Change	Global Change
cup	grasp cup with whole hand and lift across table	grasp cup with two fingers and thumb and lift across table	grasp cup with whole hand and lift higher upwards across table
lamp	reach towards touch-lamp and turn on with fingers	reach towards touch lamp and turn on with fist	arcing path of motion to reach lamp, and turn on with fingers
mug	orient towards mug, grasp handle of mug and move to self	orient towards mug and move cup towards self without grasping handle	orient head in different direction, grasp handle of mug and move to self
pencil	grasp pencil with fingers and scribble broadly on paper	grasp pencil with a closed-fist grasp and scribble	grasp pencil with fingers and scribble in different direction
drawer	orient towards drawer and push closed by grasping handle	orient towards drawer and push drawer closed without grasping handle	orient away from drawer, and push closed by grasping handle
candle	push candle across dresser with whole hand, bending forward to reach	push candle across with two fingers, bending forward to reach	push candle across with whole hand, twisting torso to reach
clay	raise hand up and hit clay with karate-chop	raise hand up and hit clay with closed fist	raise hand up higher and hit clay with karate-chop
CD	grasp CD with finger in the center, and lift across desk	grasp edge of CD, without fingers in center, and lift across desk	grasp CD with finger and lift CD higher upwards across desk

Stimuli & Materials The stimuli for Experiment 1 consisted of eight scenarios of everyday human action. Each scenario contained three videos: a standard-action video, a video altered solely in local information (local-change video), and a video altered solely in global information (global-change video). Each scenario can be viewed at: darkwing.uoregon.edu/~jloucks/actions.htm.

Local changes were alterations in the local detail of the small action performed, without changing the global spatial or properties of the action. For instance, for the standard-action video of reaching in a direct trajectory to grasp a pencil with a precision grip and then write with it, the local change was grasping with a whole hand grasp. Global changes were global modifications in the trajectory body parts followed through space. These global changes occurred without changing the small action elements. For example, for the same standard-action video of reaching and grasping a pencil, the global change was altering the trajectory of the reach to the pencil, such that the hand and arm followed an arcing motion to the pencil instead of a direct trajectory. In sum, local changes involved a substitution of a local element but retained the global aspects of the way in which the elements were executed, whereas global changes retained the individual local elements of the action but changed the way in which those elements were carried out globally.

A complete description of each action scenario is presented in Table 1. Importantly, none of the changes crossed an action/goal category boundary. In addition, the two change types did not significantly differ from one another in terms of objective, physical change (as measured by average pixel change with respect to the standard), $t(14) = 0.11$, $p > .05$.

Design & Procedure We employed a mixed design, with change type (i.e., local vs. global), and orientation (i.e., upright vs. inverted) as within-subjects variables, and presentation set as a between-subjects variable. Different presentation sets ensured no participant saw any given action scenario in both orientations. Specifically, participants in one condition viewed one set of movies inverted and the other set upright, while participants in the other condition saw the reverse. Inverted videos were rotated a full 180 degrees from their standard orientation.

On different trials, a standard video was paired with either a local change video or a global change video. The order of change videos within pairs (e.g., standard first vs. change first) was counterbalanced, yielding a total of four different trials for each video scenario in the experimental session. On same trials, a local change video, global change video, or standard video was paired with itself. The order of presentation of the pairs experienced by any given participant was randomized.

A Macintosh G4 computer was used to present stimuli and record participant responses on a 19.5" x 12" cinema display. Psychtoolbox (Brainard, 1997) was used to conduct the experiment and record responses.

After giving consent participants were seated in front of the display and the experimenter provided instructions. They were told they would be watching pairs of videos and would be asked to decide whether the videos in a pair were the same or different. Participants were informed the changes would be subtle, and that on trials when the pairs were the same the identical video file would be played twice. Participants were asked to make their judgments as quickly and accurately as possible using the computer keyboard. Participants were also told half of the videos would be presented upside down.

Results

The mean accuracy for each of the four conditions is displayed in Figure 1.[1] We predicted that accuracy for detecting global trajectory changes would be especially

[1] All means reported are accuracy for different trials only.

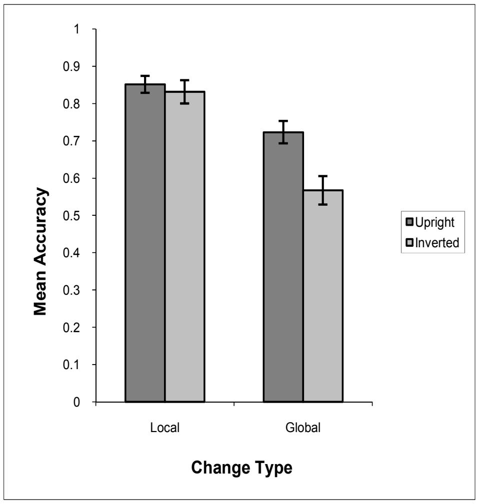

Figure 1: Mean accuracy scores as a function of type of change (local and global) and orientation (upright and inverted) in Experiment 1.

susceptible to inversion relative to accuracy for local changes. A 2 (change type) x 2 (orientation) repeated-measures ANOVA revealed a significant main effect of orientation, with upright changes (M = 78.7%, SD = 12.0%) detected significantly more accurately than inverted changes (M = 70.0%, SD = 14.7%), $F(1,33) = 16.47$, $p < .001$. There was also a significant main effect of change type, $F(1,33) = 35.09$, $p < .001$. Overall, participants were significantly more accurate at detecting local changes (M = 84.1%, SD = 13.2%) than global changes (M = 64.5%, SD = 16.9%). Finally, the predicted interaction between change type and orientation was statistically significant, $F(1,33) = 7.40$, $p < .05$.

Follow-up planned comparisons indicated that, as predicted, accuracy for detecting local changes was unaffected by inversion, $t(36) = 0.62$, $p > .05$, while accuracy for detecting global trajectory changes was significantly impaired by inversion, and $t(36) = 3.91$, $p < .001$. In fact, while detection of global changes was above chance in the upright orientation, and $t(36) = 7.45$, $p < .001$, detection of global changes was only marginally above chance when inverted, $t(36) = 1.78$, p = .084.

The fact that local changes were detected more easily than global changes was unexpected. Even in the upright format, participants were more sensitive to local changes (M = 85.1%, SD = 13.8%) over global changes (M = 72.3%, SD = 18.2%), $t(36) = 3.62$, $p < .01$. This sensitivity difference was surprising, given that the change types did not differ systematically in the amount of objective, physical change relative to the standard video. It thus appears people selectively attend to local action information over global action information.

Experiment 2

In Experiment 2, our aim was to further explore the distinction between local and global information in action. To investigate the validity of our characterization of local action information, we undertook a manipulation – low-pass filtering – expected to selectively undercut processing of local information while having little impact on processing of global information. Low-pass filtering involves removing high spatial frequencies such that only low frequency spatial information is retained. We predicted that detection of local changes should be disrupted by low-pass filtering, since these local detail changes should be dependent on high spatial frequencies, but that detection of global changes should be spared, since their detection is more heavily based on global information captured by low spatial frequencies.

Method

Participants Thirty-five adults – all University students (15 male, 20 female) – received partial course credit for their participation.

Stimuli & Materials The stimuli for Experiment 2 were the same action scenarios used in Experiment 1. Low spatial frequency versions of each video were created by rendering videos with a Gaussian blur filter via Final Cut Pro video-editing software. The blur radius was 25 and the video resolution was 720 x 480.

Design & Procedure Both the design and procedure were identical to Experiment 1, substituting low-pass filtered videos for inverted videos.

Results

The mean accuracy for each of the four conditions ispresented in Figure 2. We predicted that removal of high spatial frequencies would significantly reduce accuracy for detecting local changes, whereas detection of global

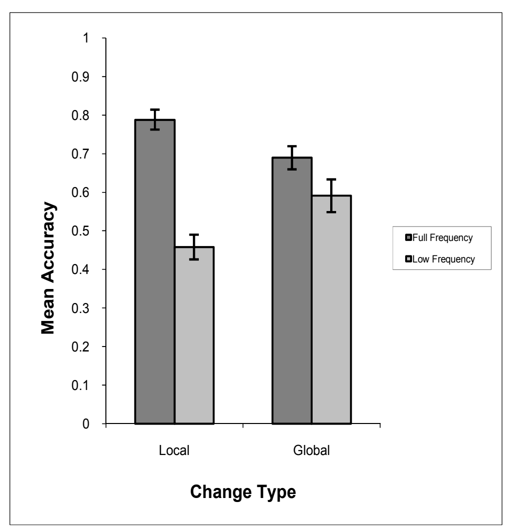

Figure 2: Mean accuracy scores as a function of type of change (local and global) and visual frequency (full and low frequency) in Experiment 2.

changes would be relatively unaffected by this manipulation. A 2 (change type) x 2 (spatial frequency) repeated-measures ANOVA revealed a significant main effect of spatial frequency, $F(1,29)= 29.24$, $p < .001$. Accuracy for low frequency videos (M = 52.5%, SD = 17.4%) was significantly reduced compared to full frequency videos (M = 73.9%, SD = 12.3%). The main effect of change type found in the previous experiment was not replicated, $F(1,29) = 0.10$, $p > .05$. However, most importantly, the analysis revealed the predicted change type by spatial frequency interaction, $F(1,29) = 15.87$, $p < .001$.

Follow-up planned comparisons indicated that, as predicted, local changes were significantly more difficult to detect when videos were low-pass filtered compared to when full frequency information was present, $t(32) = 9.37$, $p < .001$. Somewhat unexpectedly, global changes were also affected by low-pass filtering, although this effect was only marginally significant, $t(32) = 1.75$, $p = .089$. However, accuracy for local change detection in the low frequency format was so low it was not significantly different from chance, $t(32) = -1.30$, $p > .05$, while accuracy for global changes in the low frequency format remained above chance, $t(32) = 2.15$, $p < .05$.

Participants' heightened sensitivity to local over global information was again present in this experiment. Participants were significantly more sensitive to local information in the full frequency format (M = 78.8%, SD = 14.8%) than global information in the full frequency format (M = 68.9%, SD = 17.2%), $t(32) = 2.74$, $p < .05$.

General Discussion

While any action can be analyzed along a computationally immense number of dimensions, it is unknown which dimensions people actually pay attention to when observing human action. On analogy with face-processing research, we examined the possibility that observers may key on highly relevant local detail (local information) as well as more global motion information (global information) for the purposes of action discrimination. We found that observers are sensitive to both sources of information for this purpose. At the same time, our findings provide clear evidence that local action information is elevated in adults' processing relative to global action information. People more readily detect changes to local detail across the board – regardless of whether action is viewed upright or inverted – and this is so even when local and global changes are equivalent in terms of actual objective, physical change. We also found that processing of global information is especially reliant on viewing action in the upright format. Local information, on the other hand, is especially reliant on high spatial frequencies for accurate processing. These findings are consistent with how local and global information are processed in the face, and thus there are striking similarities in how these two sources of information manifest in processing across the two domains. These results are the first of their kind highlighting the sources of information people readily make use of when discriminating dynamic human actions.

Selective attention to local information in action

In our experiments, changes in local detail within action impacted participants' detection of action differences to a markedly greater degree than changes in global characteristics of the action. We were able to confirm that differences in the magnitude of physical motion change were not the cause of this heightened sensitivity to local information. Thus, the local and global changes in our stimuli did not differ in physical magnitude, yet they differed noticeably in *psychological magnitude* for our participants.

Recent research investigating eye gaze during action viewing meshes nicely with this selective attention effect. Flanagan and Johansson (2003) have demonstrated that when people view others performing object-directed actions, their eye movements "predict" the goals of the actor's arm trajectories. In our stimuli, local changes could largely be characterized as changes in hand-object relationships. If predictive gaze shifts to the region of contact between the hand and the object are automatic, then participants would have had little trouble detecting changes in this region. Such automaticity could also help to explain their relative difficulty detecting global changes, as the global changes in our stimuli tended to occur outside of the hand-object contact region. All in all, this eye-gaze research and our own findings provide converging evidence for selective attention to local detail in action processing.

The local/global distinction in action

Another important distinction between local and global action information is that their processing depends upon distinct stimulus properties. Processing of local information relies heavily on the identification of local detail registered in high spatial frequencies, while global information appears to trade more heavily on a standard, upright action template. Our conceptualization of this template is similar to the template proposed for the human body (Reed et al., 2003), but also includes a dynamic component. Although previous research has demonstrated processing is disrupted when action is inverted (e.g., Shipley, 2003), the current research clarifies that this is largely due to the fact that processing global information is severely undercut by inversion.

A domain-general dual processing approach?

As we have shown with these experiments, local and global information are processed similarly in the face and in human action. Although rarely discussed explicitly, careful examination of findings in the face processing literature reveals a selective attention effect for local information that parallels what we have seen in the action domain.

Our findings lend support to the idea that a local/global processing distinction is not unique to faces, and may be a domain general processing approach that develops and is utilized whenever stimuli are processed with expertise (Diamond & Carey, 1986; Gauthier & Tarr, 1997). As we observe action, at times we must be sensitive to subtle properties of action in order to make correct inferences about intentions, situational factors, or to predict future behavior. As in the face, identification and discrimination at such a fine level seems to demand a powerful processing approach. Sensitivity to two distinct sources of information, processed in parallel, could be an efficacious solution. In partial support of this view, evidence suggests similar brain regions implicated in processing local versus global information in the face are also involved in expert object identification (Rossion et al., 2002). Although not critical to supporting this domain generality view, it would be interesting to explore whether brain regions used in the local/global distinction in face processing and expert object identification are also used in processing these distinct sources in human action.

Future investigations

Although these experiments highlight the basic role of local and global sources of information in action discrimination, a number of unanswered questions arise. First, it would be of interest to pursue converging evidence that local information is selectively attended to relative to global information. For example, if local information is elevated in people's processing then there might be little detriment to detection of local changes in a dual-task situation, while detection of global changes might suffer more drastically.

Another question concerns how entrenched the selective advantage for processing local detail is in people's processing. It is unknown whether it can be shifted with explicit or contextual manipulations that shift attention to global properties of action.

The developmental origin of this dual-processing approach is also ripe for investigation. For face processing, evidence suggests processing of global information develops more slowly than processing of local information (Mondloch et al., 2002). Whether a similar developmental pattern holds for human action is of significant interest.

Conclusion

These experiments indicate adults track both local and global action information for identification and discrimination. Psychologically speaking, changes in local action detail loom large in adults' processing relative to global properties. Moreover, processing of local versus global information can be impacted in contrasting ways with different manipulations. This pattern of findings directly parallels those documented in adults' processing of faces, suggesting the local/global dual-source processing approach observed in these two domains may arise through domain-general mechanisms sub-serving observational expertise.

Acknowledgements

This research was funded by NSF Grant No. BCS-0214484, awarded to the second author. We are also grateful to Bridget Klest for providing helpful comments on a previous draft of this paper.

References

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433-436.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*, 107-117.

Flanagan, J., & Johansson, R. S. (2003). Action plans used in action observation. *Nature, 424*, 769-771.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception, 29*, 159-170.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*, 1673-1682.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception, 34*, 77-86.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*, 201-211.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*, 255-260.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception, 31*, 553-566.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality & Social Psychology, 28*, 28-38.

Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological Science, 14*, 302-308.

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychological Science, 13*, 250-257.

Shipley, T. (2003). The effect of object and event orientation on perception of biological motion. *Psychological Science, 14*, 377-380.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141-145.