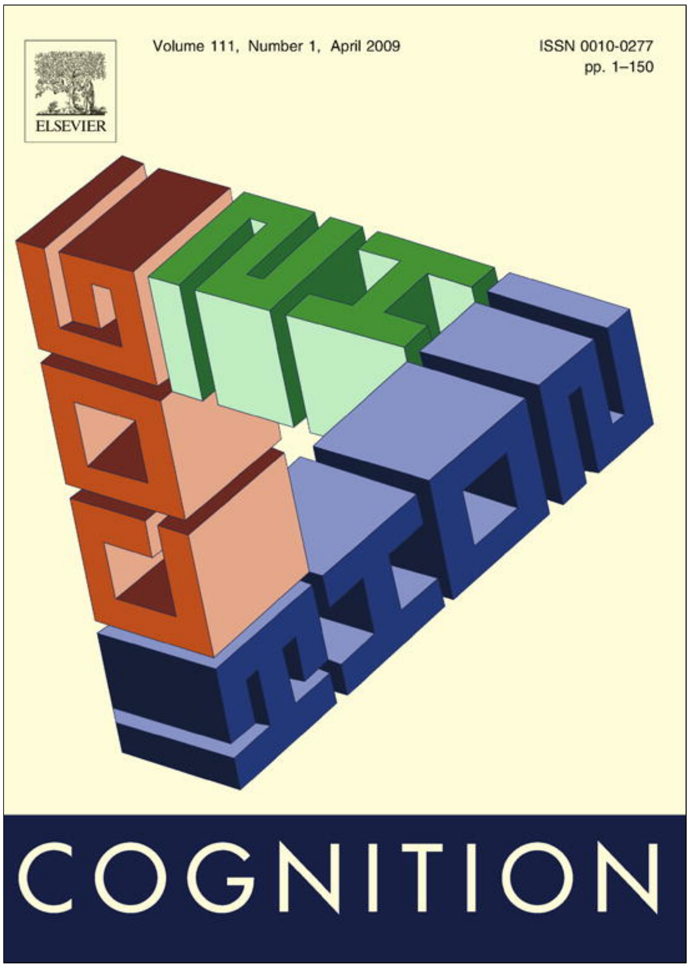
Volume 111, Number 1, April 2009
ISSN 0010-0277
pp. 1–150
ELSEVIER
COGNITION

Cognition 111 (2009) 84–97

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Sources of information for discriminating dynamic human actions

Jeff Loucks *, Dare Baldwin

Department of Psychology, 1127 University of Oregon, Eugene, OR 97403-1227, United States

ARTICLE INFO

Article history:
Received 8 January 2008
Revised 17 December 2008
Accepted 30 December 2008

Keywords:
Human action
Face processing
Intentions
Selective attention
Discrimination
Information processing

ABSTRACT

Despite the importance of action identification and discrimination in action perception and social cognition more broadly, little research has investigated how these processes are achieved. To this end, we sought to identify the extent to which adults capitalize on featural versus configural sources of information when discriminating small-scale human actions such as grasp and place. Results across two experiments indicate adults are sensitive to both sources of information in action discrimination, but selectively attend to featural over configural action information. The findings also parallel what is known regarding face processing: processing of configural information is especially disrupted by inversion, whereas processing of featural information is specifically affected by low-pass filtering.

1. Introduction

Every day we are witness to the exceptionally complex stimulus that is other people's behavior. People produce a wide variety of actions, often assembled in novel sequences, involving contact with many different objects in rapid succession. Action is typically fluid and largely continuous, and virtually no motion stream entirely replicates what we have previously encountered. Occasionally, we are even witness to radically novel actions performed on new objects. 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, however, essential for everyday functioning and social interaction. Whether it's preparing dinner with a partner or navigating through a crowded airport, success is achieved, in part, by understanding others intentions via their actions.

Given the centrality of action processing to everyday cognitive and social functioning, surprisingly little is known about the fundamental processes that are involved. Seminal work by Newtson and colleagues (Newtson, 1973; Newtson & Engquist, 1976) as well as Johansson (1973) provided some of the initial groundwork for research on action segmentation and perception. In recent years research in this domain has grown considerably, from investigation of neural circuitry involved in action processing (Rizzolatti & Craighero, 2004; Zacks et al., 2001), to hierarchical processing of action (Hard, Lozano, & Tversky, 2006), and segmentation of human action by adults (Baldwin & Baird, 2001; Zacks, Tversky, & Iyer, 2001) and young infants (Baldwin, Baird, Saylor, & Clark, 2001; Saylor, Baldwin, Baird, & LaBounty, 2007).

However, there has been scant research on how people approach the fundamental task of identifying and discriminating human actions. Discrimination is one component of the basic process of action categorization (the other being generalization). Although every instance of human action is a unique event (exempting video replay), it seems clear that actions are treated as category members in our processing of the behavior stream. That is, members of a kind are processed as similar despite the diversity of objects targeted by action and considerable variability in the rates and trajectories of motion across individual instances. Rapidly discriminating actions on

* Corresponding author. Tel.: +1 541 346 2086; fax: +1 541 346 4911.
E-mail address: jloucks@uoregon.edu (J. Loucks).

0010-0277/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2008.12.010

the category level is a skill employed countless times per day in everyday life. For example, subtle discriminations aid in knowing whether the individual at the coffee shop is waiting in line or just milling about, or whether one's breakfast partner is reaching out with a gesture of endearment or instead is just reaching for the jam. Moreover, imitation – a hallmark of human social cognition – requires an analysis of the behavior stream finely-tuned to discriminate between different functional actions (Meltzoff, 2002).

One body of research that has provided information relevant to how people identify and discriminate actions is that of biological motion processing. Videos of point-light action – in which actors are filmed in the dark, with lights affixed to their major joints – have been used to explore low-level visual mechanisms that support the identification of biological versus non-biological motion (Johansson, 1973). This research indicates that people can use relational information, in the form of the relative motion of the joints, in order to identify and discriminate some broad types of actions (e.g., walking, running). Developmental work has ascertained that even young infants can discriminate relational aspects of motion present in some point-light displays (Bertenthal, Proffitt, & Cutting, 1984), and that young children can extend motion verbs (e.g., running) to such displays (Golinkoff et al., 2002). Thus, such low-level relational information appears to be sufficient for identifying and discriminating some basic forms of human action.

Although research using point-light displays supports the idea that relational information can be used to identify and discriminate point-light displays of action, this body of research has been less informative about people's *everyday* action processing skills. To date, most point-light research has used a restricted range of simplified action stimuli (iterative and transitive whole body actions). Processing everyday action likely requires more than low-level relational information, 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. However, little research to date has addressed this issue.

Our goal with the current research was to explore basic kinds of information people might extract from the motion stream to assist in identifying and discriminating between everyday, dynamic human actions. What kind of information do people key on, for example, in judging an action to be a slap versus a caress? Although one could readily analyze the physical motion differences between these two actions, the number of possible points of comparison between the two 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 focused heavily on the same question regarding sources of information utilized in identification and discrimination – the domain of face processing.

2. Faces and action

Faces and action have much in common. Like action, processing of faces seems, at least phenomenologically, to require minimal effort: we have little trouble discriminating the faces of good friends from those of mere acquaintances, despite variations in visual conditions and the demand for high accuracy. In addition, efficient processing of faces, like action, has important consequences for social cognition.

People are known to be sensitive to at least two sources of information in faces: salient featural information, and more global, configural information (Maurer, Le Grand, & Mondloch, 2002).[1] Featural information has been characterized as information about highly defined local elements of the face, such as the eyes, nose and mouth. In contrast, configural information has been characterized as more widely distributed information about the spatial relations among these featural elements.[2] Although featural information intuitively captures the salient elements of the face, an increased sensitivity to configural information in the face, compared to other objects, is believed to be one facet of expertise in face recognition (Valentine, 1988).

Given the similarities between faces and human action, we questioned whether they might also share analogous processing characteristics, and in particular, whether featural and configural sources of visual information might operate in human action processing as in face processing. Some researchers have argued that this dual processing approach might not be specific to faces, and may develop whenever stimuli are processed with expertise (Diamond & Carey, 1986; Gauthier & Tarr, 1997; Valentine, 1988). Furthermore, since these two sources of information are often correlated with spatial frequency, we also felt it was important to investigate their processing in action, as spatial frequencies are a fundamental component of all visual processing (e.g., Morrison & Schyns, 2001).

In the face, evidence suggests featural versus configural sources of information are relatively distinct from one another. One such source of evidence is the "face-inversion effect" – the finding that faces are difficult to recognize when presented upside-down (e.g., Bartlett & Searcy, 1993; Collishaw & Hole, 2002; Hole, George, Eaves, & Rasek, 2002; Leder & Bruce, 2000; Valentine, 1988; Yin, 1969). The general consensus within this literature is that inversion specifically disrupts the processing of configural information in the face while having relatively little impact on the processing of featural information (Freire, Lee, & Symons, 2000; Mondloch, Le Grand, & Maurer, 2002; Murray, Yong, & Rhodes, 2000). Presumably the standard configural face template people are accustomed to viewing in the upright format (eyes above nose, nose above mouth) is less

[1] Featural information is sometimes referred to as "local" information, and configural information is referred to as "global" information. We have decided to use what we feel are more descriptive terms, which best capture the distinction between these two sources in the face.

[2] Our use of the term configural information here refers to what is also known as second-order configural information. Configural information in the face has been defined in many ways: see Maurer et al. (2002) for a review.

accessible in the inverted stimulus, and thus recognition and discrimination suffer.

Additional evidence that these two sources of information in the face are separable comes from demonstrations that their processing is reliant on relatively distinct high and low spatial frequencies. High spatial frequencies (measured in cycles per degree of visual angle) encode information about local details in an image, on which identification of facial elements is thought to depend. Low-pass filtering removes high spatial frequencies, and thus impacts face processing in a manner directly opposite to the effect of inversion – primarily disrupting featural processing while leaving configural processing relatively unaffected. Although this idea was first proposed over two decades ago (Sergent, 1986), evidence that featural information in the face relies on high spatial frequencies has been documented only recently (Goffaux, Hault, Michel, Vuong, & Rossion, 2005). Goffaux et al. showed that discrimination based on featural information suffers at low spatial frequencies, while discrimination based on configural information is relatively unaffected.

Although not often discussed outright, evidence also indicates that people selectively attend to featural information over configural information in the face. Even when faces are presented in their normal orientation, at full visual frequency, people detect changes in featural information at much higher rates than they do for changes in configural information. In fact, in order to equate detection accuracy between the two types, one must select configural changes that are objectively larger visual changes (in terms of the amount of pixel change) than the featural changes (Yovel & Kanwisher, 2004, also see Mondloch et al. (2002) for a similar point). This dominance of featural information over configural information may be a general difference between these two sources of information, or it may be dependent on the context of the discrimination task (see Morrison & Schyns, 2001, for a discussion of this issue). Importantly, although researchers often overlook this attentional difference, we believe it is an important part of the distinction between these two sources of visual information.

Finally, neuropsychological evidence corroborates the claim that there is a meaningful distinction between these two sources of information, with demonstrations that processing of these two sources show distinguishable neural correlates (Rossion et al., 2000) and differential hemispheric specialization (Hillger & Koenig, 1991). Moreover, early visual deprivation in the right hemisphere affects featural processing but not configural processing (Le Grand, Mondloch, Maurer, & Brent, 2003).

To sum up, a wealth of research indicates that featural and configural aspects of the face are distinct sources of information for the purposes of recognition and discrimination. If processing action relies on mechanisms similar or related to those involved in face processing, then dynamic analogues of featural and configural information in action might be meaningfully distinguishable from one another, and sensitive to manipulations such as orientation changes and low-pass filtering in ways that parallel sensitivity in face processing.

A small amount of existing empirical evidence helps to increase the plausibility of our contention that a distinction between featural and configural information operates in action processing much like what has been documented in face processing. First, processing of the human body appears to be sensitive to orientation (Reed, Stone, Bozova, & Tanaka, 2003). Reed et al. had participants detect configural body posture changes in static pictures of the human body in upright and inverted orientations. Participants were significantly less accurate and slower at detecting body posture changes when they were presented inverted as opposed to upright. Second, inversion also disrupts biological motion processing (Pavolva & Sokolov, 2000; Shipley, 2003). Importantly, Shipley demonstrated that the processing of dynamic motion information actually is disrupted when biological motion is inverted, and it is not simply that the processing of the inverted human body gives rise to this effect. Participants in this task detected whether a point-light individual, either walking on feet or on hands, was present amidst a certain degree of additional point-light noise. Thus, in the upright orientation the individual was "upside-down" but the motion was upright, and in the inverted orientation the individual was "upright" but the motion was inverted. Participants were more skilled at detecting the walking in upright conditions, regardless of the orientation of the individual. This suggests it is inversion of human motion per se that gives rise to difficulties in processing action presented upside down.

All in all, it seemed a credible possibility that action processing may utilize distinct sources of featural and configural action information, similar to the analogous sources of information used in face processing. But before exploring these issues, we must first spell out the form featural and configural information might take in the action context.

3. Featural and configural information in action

Although action is inherently dynamic and relational in nature, certain dynamic aspects of the motion stream seem to be critical in people's judgments about the identity of an action, and these highly relevant characteristics seem often to be featural in nature. For example, identifying an action as a grasp relies on identifying dynamic motion that displays 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 of a grasp. When detailed, relatively 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 of featural information in action is consonant with the characterization of featural 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 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 changes could be made in the trajectory of a reach toward an object without affecting the judgment that this was a grasping action. However, the spatial trajectory is a crucial component of the grasping action, and alterations in this dynamic property do inform goal identification: processing the trajectory of a grasp is key to identifying the specific goal of that grasp. Configural action information thus represents the way in which featural action elements are spatially related to one another, much as configural information has been conceptualized in the face.

For purposes of this research, then, we defined featural information in action as local detail surrounding 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 action (Baird, Baldwin, & Malle, 2008; Baldwin, 2005; Hard et al., 2006; Newtson, 1973; Zacks et al., 2001). 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 through the air, and then placing it down in a new location. Identifying the small-scale action unit of grasping requires relatively local detail about a particular hand shape in relation to the object. Put another way, we conceptualized featural information as highly relevant local detail within the motion flow that observers might plausibly rely on to identify instances of small-scale actions and differentiate them from instances of other small-actions. Configural action information, in contrast, was conceptualized as global relational properties of the motion stream manifested across one or more small-scale actions, such as the overarching trajectory of motion. Pursuing the book-moving example, configural information might be information about what path the hand took through space on its way to the book.

It is important to note that the distinction between featural and configural information is not all-or-none, but is instead continuous in nature. Aspects of the visual scene that are segmented as features can be re-analyzed in terms of spatial relations within the feature, and aspects that are not typically analyzed as features can be processed in a piecemeal fashion. This is true in the face, and is also true in human action. What makes them distinct is that people *typically* process some aspects as features and other aspects as configurations. The continuous nature of this distinction is crucial for understanding that in terms of spatial frequency these two sources of information exist on a gradient. Within any action event, dynamic visual information is captured across a large continuum of spatial frequencies. We expect featural and configural action information to be represented at relatively disparate ends of the spatial frequency continuum – featural information at relatively high spatial frequencies, and configural information at relatively low spatial frequencies. Relatedly, there are no particular spatial frequencies that are invariably associated with featural and configural information in action: spatial frequencies for these two sources of information are dependent on the distance and the viewing angle of the observer relative to the action event. However, featural information will always be represented at higher spatial frequencies relative to configural information.

To investigate the meaningfulness of the featural versus configural 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 (in sequence) depicting everyday action and made same/different judgments about these pairs. On *different* trials, videos with alterations in either featural information or configural information (but not both) were paired against the standard action video. Featural change videos were identical to the standard video except for a single alteration in the dynamic local detail of a small-scale action in the motion stream. Configural change videos were also identical to the standard video except for a single change in the global spatial properties of the motion stream. In Experiment 1, we investigated whether processing of configural action information is especially reliant on viewing motion in an upright orientation. In Experiment 2, we utilized low-pass filtered versions of the same videos to explore whether featural and configural action information are processed using relatively distinct spatial frequencies.

A critical issue of concern for these experiments was measuring how different our featural and configural alterations were in the amount of objective, physical change to the action with respect to their standard videos. As mentioned earlier, previous face processing research documents that people are more sensitive to featural facial changes compared to configural facial changes. In our research, we wanted to equate sensitivity to featural and configural action information in the upright/full frequency format, so that interpreting differences related to our manipulations (inversion and low-pass filtering) would not be clouded by performance differences in the upright/full frequency format. However, we also wanted to explore whether a similar selective attention effect may exist in human action for featural information over configural information. Thus, in our experiments we purposely tried to make configural changes *larger* than featural changes, so that even if participants were more sensitive to featural changes over configural changes, the processing disparity would be reduced. Thus, we aimed to give configural changes an advantage over featural changes in terms of detection on the basis of objective, physical change.

4. Experiment 1

In Experiment 1, we investigated whether an inversion effect might selectively impair processing of configural information in action. One can imagine that processing configural information might trade heavily on a standard prototypical template for human action (head motion above the arms, leg motion below the waist, etc.), similar to the prototypical templates have been implicated in face and body processing. If so, inverting action might radically undercut the ability to track configural changes in action, as it does in face processing.

We predicted that observers would be sensitive to both featural and configural changes when action was presented in its normal, upright orientation. However, we predicted that inversion of video pairs would disrupt change-detection sensitivity for pairs differing solely in configural information to a significantly greater extent than pairs differing solely in featural information.

4.1. Method

4.1.1. Participants

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

4.1.2. Stimuli and Materials

The stimuli for Experiment 1 consisted of eight different scenarios of everyday human action, performed by a single male actor. Three different videos were created for each action scenario: a standard-action video, a video altered solely in featural information (featural change video), and a video altered solely in configural information (configural change video). All videos were recorded digitally without sound. The videos ranged in length from 2.2 to 5.13 s ($M = 4.26$, $SD = 0.89$), and the frame rate was 30 frames per second. Each video measured 640 (w) $\times$ 480 (h) pixels.

Still frames depicting the standard, featural, and configural change videos from one video set appear in Fig. 1. Featural changes were modifications in the local detail of the small action performed, without altering the global spatial trajectories of body parts. For instance, for the standard-action video of reaching and grasping a cup with a whole-hand grasp and moving it in a direct trajectory across the table, the featural change was grasping the cup with a pincher-grasp instead. Configural changes were global modifications in the spatial trajectory body parts followed through space, without altering the small action detail. For example, for the same standard-action video of reaching and grasping a cup with a whole-hand grasp and moving it in a direct trajectory across the table, the configural change was altering the trajectory of the relocation, such that the hand and arm lifted the cup higher off the table in its movement to the new location instead. Configural changes were not limited to arm trajectory, however, and included modifications of the head and torso trajectories also – all configural in nature. In sum, featural changes involved an alteration in local action detail but retained the broader spatial aspects of the way in which the action was executed, whereas configural changes retained the local detail of the action but altered the broader spatial elements of the action.

A complete description of each video scenario, including the standard, featural change, and configural change videos, is presented in Table 1. As is clear from the table, the eight scenarios chosen for the stimulus set depicted a range of different object-oriented actions enacted in relatively diverse indoor settings (e.g., kitchen, bedroom, office, etc.). Inverted versions of each video were created using Final Cut Pro video editing software (see Fig. 2).

To the extent possible with live action, a concerted effort was made to ensure that both featural and configural changes altered only one particular aspect of the motion, while all other dimensions remained unchanged. For example, if the configural change was in the trajectory the arm took to reach an object, then only the trajectory of the arm was changed, and elements such as head orientation and type of hand contact remained unchanged. As well, the initial actor and object locations, lighting, facial expression, and distance from the viewer were always equated across videos within a set. Also note that both featural and configural changes represented within-category changes to the action. Neither featural changes nor configural changes altered the outcome of the action, nor did either change cross a verbal category (e.g., no change would require the use of a different action verb to describe the action).

As stated earlier, another issue of concern was measuring the amount of physical motion change between featural and configural change videos with respect to the standard videos. For each change video, we calculated the overall physical amount of change between a change video and its respective standard video using an algorithm that compared the amount of pixel change between each frame in the change video and each frame in the standard video. Individual frames from a change video were temporally aligned with the individual frames of its standard video (e.g., the first frame of the standard video aligned with the first frame of the featural change video, the second standard frame aligned with the second configural change frame, etc.), and the amount of pixel change between these aligned frames was calculated using the following algorithm:

$$\sum_{i=1}^{h}\sum_{j=1}^{w}\sqrt{(R_{Cij}-R_{Sij})^2+(G_{Cij}-G_{Sij})^2+(B_{Cij}-B_{Sij})^2},$$

where R, G, and B represent the red, green, and blue color values of a pixel, C and S denote the change frame and standard frame, i and j represent the coordinate value of the pixel, and h and w represent the height and width of the frames in pixels.

The average physical change across a video was then calculated for each change video/standard video comparison. The rationale behind this comparison was that, given that the videos were all filmed on a tripod from a single visual angle and the actor's starting position was virtually identical at the outset of all videos, change in pixels between two aligned frames would be due to motion-driven position changes in the actor's body and/or the objects he was acting on. Thus, motion change between two frames would be small, for instance, if the actor's arm position in the change video frame was similar to its position in the standard video frame and overlapped to a high degree, compared to when it was farther away from its position in the standard video frame and the overlap was reduced.

As we had planned, the pixel-change test demonstrated that configural change videos ($M = 2{,}879{,}811$, $SD = 713{,}200$) were significantly larger physical motion changes compared to featural change videos ($M = 2{,}667{,}167$, $SD = 676{,}715$), $t(7) = 2.64$, $p = .03$, $d = 0.93$. Thus, the two types of changes were not equated in an objective sense. Rather, configural changes had an objective advantage in the magnitude of physical change compared to featural changes.

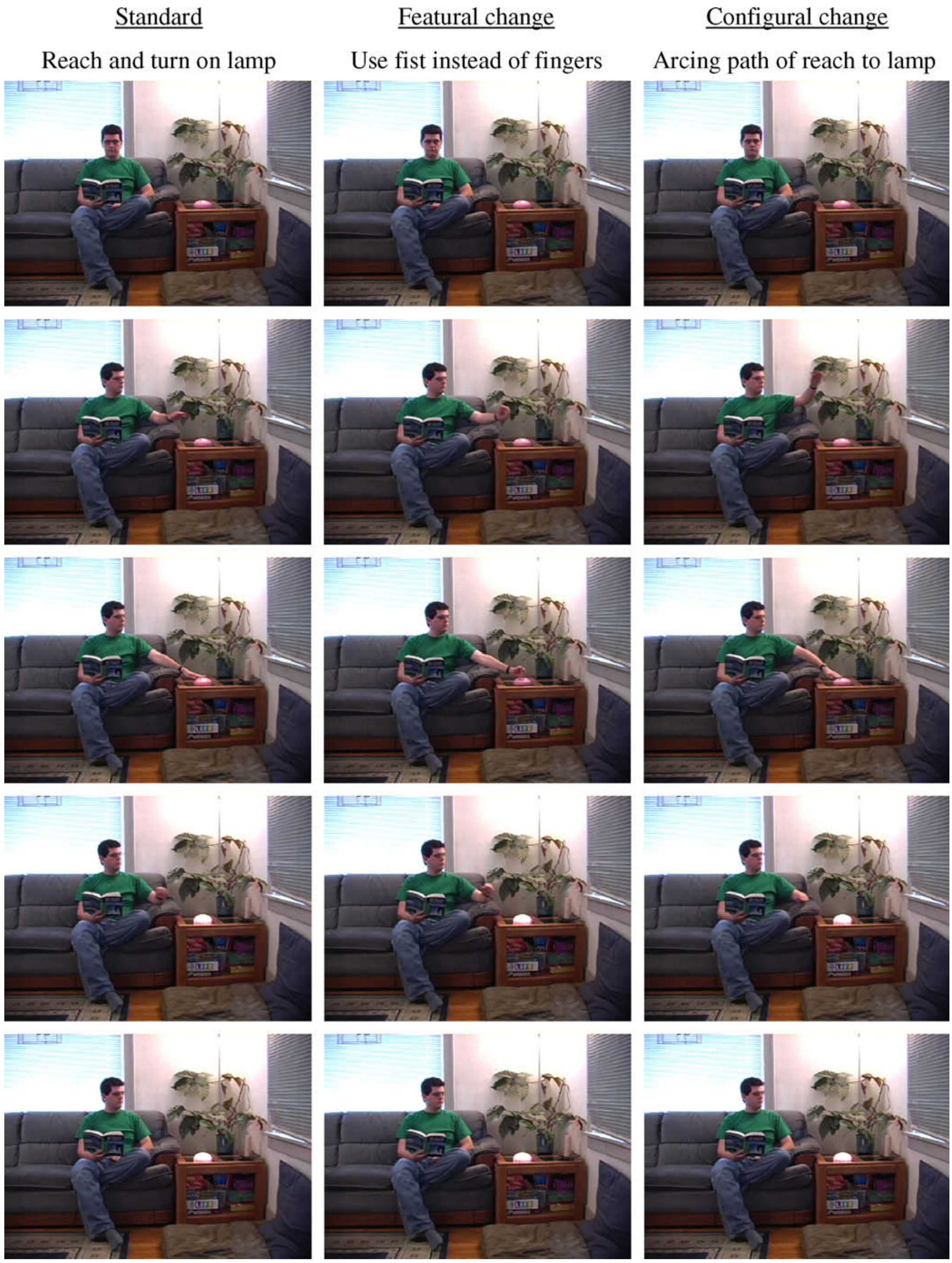

Fig. 1. An example of a featural and configural action change (the *Light* scenario).

4.1.3. Design and procedure

We employed a mixed design, with change type (i.e., featural versus configural change), and orientation (i.e., upright versus 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 the two presentation set groups viewed different sets of upright and inverted action scenarios. Thus, participants in one condition viewed set A (candle, clay, drawer, and lamp) inverted, and set B (cup, cd, mug, and pencil) upright. Participants in the other condition saw the reverse: set B inverted and set A upright. Inverted videos were videos rotated a full 180° from their standard orientation. This 2 (change type) × 2 (orientation) design engendered four types of trials for all participants.

On *different* trials, a standard video was paired with either a featural change video or a configural change video. The order of videos within pairs (e.g., standard video

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

Scenario	Standard video	Featural change	Configural 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

first versus change video first) was counterbalanced, yielding a total of four *different* trials for each video scenario in the experimental session. On *same* trials, a standard video, a featural change video, or a configural change video was paired with itself (each standard video was paired with itself twice in order to equate the number of *same* and *different* trials in the experimental session). This design yielded a total of 64 trials: 32 upright (16 same, 16 different), and 32 inverted. Thus, participants saw a video from each of the action scenarios a total of 16 times. The order of presentation of the pairs experienced by any given participant was randomized.

A Macintosh G5 computer was used to present stimuli and record participant responses on a 19.5′ × 12′ cinema display. From where participants were seated, videos subtended ~9.29° of visual angle. 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 informed that 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 that the changes would be subtle, and that on trials

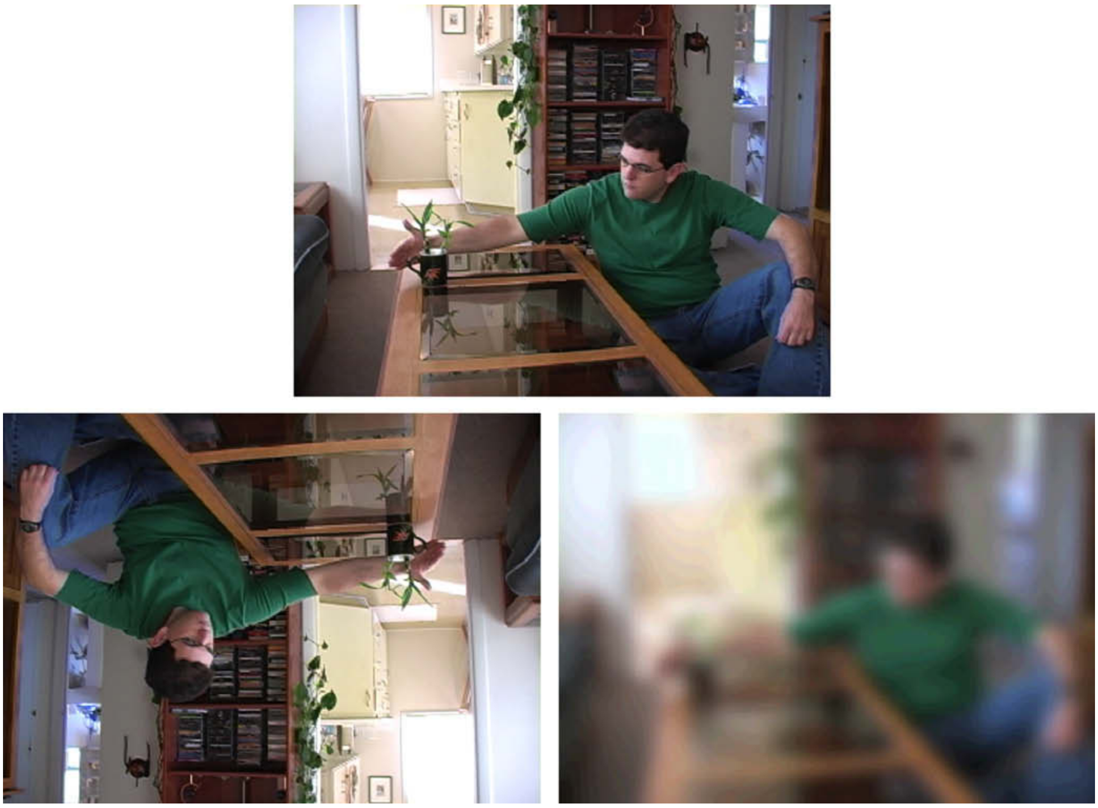

Fig. 2. Examples of a standard video (top), of the inversion manipulation used in Experiment 1, (bottom left), and of the low-pass filtering manipulation used in Experiment 2 (bottom right).

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 informed of the fact that half of the video pairs they would be seeing would be presented upside down. We requested that they not tilt their heads to the side while viewing the inverted pairs.

4.2. Results and discussion

In order to assess participants' sensitivity to each of the change types in each condition, accuracy data were converted to d' scores for all statistical analyses.

The mean d' scores for each of the four conditions appear in Fig. 3a. We predicted that sensitivity to detecting configural trajectory changes would be especially susceptible to inversion relative to sensitivity to detecting featural changes. A 2 (change type) × 2 (orientation) repeated-measures ANOVA revealed a significant main effect of orientation, with upright changes ($M = 1.91$, $SD = 0.50$) detected significantly better than inverted changes ($M = 1.72$, $SD = 0.55$), $F(1,35) = 4.77$, $p = .04$, $\eta_p^2 = .12$. There was also a significant main effect of change type, $F(1,35) = 39.23$, $p < .001$, $\eta_p^2 = .53$. Overall, participants were significantly more sensitive to featural changes ($M = 2.15$, $SD = 0.61$) than configural changes ($M = 1.50$, $SD = 0.50$). Finally, the predicted interaction between change type and orientation was statistically significant, $F(1,35) = 4.89$, $p = .03$, $\eta_p^2 = .12$.

Follow-up planned comparisons indicated that, as predicted, sensitivity to featural changes was unaffected by inversion, $t(38) = -0.21$, $p = .83$, while sensitivity to configural trajectory changes was significantly impaired by inversion, $t(38) = 2.79$, $p = .008$, $d = 0.45$.

Despite the fact that we found configural changes to be significantly larger than featural changes in terms of objective, physical differences, participants displayed heightened sensitivity to featural action information compared to configural action information. Even when comparing sensitivity in the upright format alone, participants were significantly more sensitive to featural changes in the upright format ($M = 2.14$, $SD = 0.66$) than configural changes in the upright format ($M = 1.69$, $SD = 0.59$), $t(38) = 3.81$, $p < .001$, $d = 0.61$.

The fact that participants were less sensitive to configural changes compared to featural changes, even in the upright format, may have influenced the interaction between change type and orientation. If configural changes are generally less detectable than featural changes, then perhaps configural change detection was more affected by inversion simply due to inherent fragility of processing. To test this possibility, we analyzed the correlation between each participant's mean sensitivity to upright configural changes and each participant's mean sensitivity to inverted configural changes.[3] If overall processing fragility were the primary source of participants' difficulty detecting configural changes in the inverted format, then one would predict a positive correlation between these two indices. However, this correlation was not significant, $r(39) = .16$, $p = .34$. This finding thus provides some evidence suggesting that the observed interaction between change type and orientation was not simply a function of the fragility in processing of configural information overall.

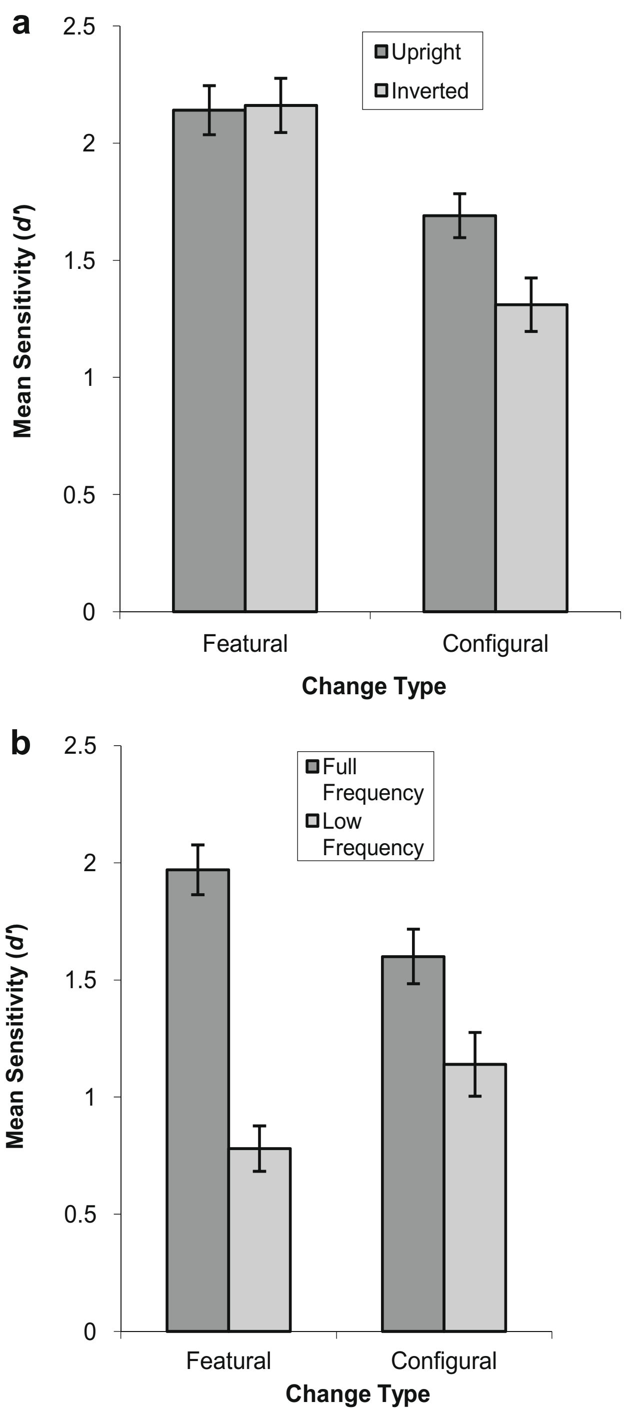

Fig. 3. Mean d' scores as a function of (a) change type (featural and configural) and orientation (upright and inverted) in Experiment 1, and (b) change type and visual frequency (full and low frequency) in Experiment 2 (error bars represent standard error of the mean).

[3] These two variables are independent of one another, due to the fact that participants saw different videos upright and inverted.

The results of Experiment 1 clearly indicated that people are sensitive to both featural and configural information in action, and also suggested that there is a meaningful distinction to be made between these two sources of information. First, the findings revealed some evidence suggesting that adults selectively attend to featural action information relative to configural action information: in the upright format adults displayed greater sensitivity to featural changes over configural changes, despite the fact that the magnitude of physical change was significantly *greater* for configural changes. Put another way, adults generally discriminated actions differing in configural information, but these substantial configural differences were less salient relative to much smaller featural changes. Interestingly, as mentioned earlier, this same pattern of greater sensitivity to featural information is often apparent in the face processing literature (Freire et al., 2000; Goffaux et al., 2005; Leder & Bruce, 2000; Mondloch et al., 2002), but is frequently either overlooked or given little consideration. In the present study, participants were not given any information about what kind of changes they would be seeing, so this selective bias to attend to featural information in human action appears to be default or implicit.

Second, these findings point to configural information within action being more readily disrupted by inversion than featural information. When the videos were presented upside down, processing of configural information suffered significantly, while processing of featural information was spared. Additional analyses provide no evidence that this differential sensitivity to orientation was related to the fact that configural changes were harder to detect overall. Together, the package of findings suggests that processing of configural information heavily relies on a prototypical action template, with the head motion above the arms and the arm motion above the trunk, acting in accord with standard physical principles (e.g., direction of gravitational force). This finding meshes nicely with the findings of Shipley (2003), and demonstrates further that inversion specifically undercuts the processing of configural information, even for everyday human actions presented in perceptually rich contexts.

The findings of Experiment 1 were also consonant with the literature regarding analogous featural versus configural sources of information in the face. In the face, the processing of configural spatial-relational information is especially reliant on upright processing, while featural information is not (Freire et al., 2000). Thus, these two sources of information across these different domains appear to share processing similarities.

An issue as yet unresolved by the research was whether featural and configural changes are reliant on relatively distinct high and low spatial frequencies, as their facial analogues are. Another important question was whether adults' detection of featural changes could be selectively impaired relative to detection of configural changes. To the extent that featural and configural action information are processed as unique sources within dynamic action, we might also expect people to respond to them differently depending on the type of manipulation. Experiment 2 addressed these two issues.

5. Experiment 2

In Experiment 2, our aim was to further explore the distinction between featural and configural action information in the upright format. Recall that, on analogy with the face processing literature, we characterized featural information in action as dynamic information that is captured in local detail surrounding a small action element. In face processing, analysis of facial features in this piecemeal fashion led researchers to hypothesize that featural processing relies on relatively high spatial frequencies in comparison to configural processing. Research using low-pass filtering – removing high spatial frequencies such that only low-frequency spatial information is retained – has indeed documented that featural processing relies on high spatial frequencies to a much greater degree than does configural processing (Goffaux et al., 2005). If featural action information is processed in similar fashion to featural processing in the face, then we also expect this source of information to be similarly reliant on high spatial frequencies. Thus, in Experiment 2, we used low-pass filtering to investigate the role of spatial frequency in the processing of these two sources of information in human action. We predicted that detection of featural changes should be disrupted by low-pass filtering, since these detailed changes should be dependent on relatively higher spatial frequencies, but that detection of configural changes should be spared, since their detection is more heavily based on information captured by relatively lower spatial frequencies.

5.1. Method

5.1.1. Participants

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

5.1.2. Stimuli and Materials

The stimuli for Experiment 2 were the same eight action scenarios used in Experiment 1. Low spatial frequency versions of each video were created by rendering videos with a Gaussian blur filter (pixel radius of 25) via Final Cut Pro video-editing software (see Fig. 2 for an example).

5.1.3. Design and procedure

Both the design and procedure were identical to Experiment 1 except in one respect: participants were told that half of the videos would be blurred (as opposed to inverted).

5.2. Results and discussion

The mean d' scores for each of the four conditions are displayed in Fig. 3b. We predicted that removal of high spatial frequencies would significantly reduce accuracy for detecting featural changes, whereas detection of configural changes would be unaffected by this manipulation. A 2 (change type) $\times$ 2 (spatial frequency) repeated-measures ANOVA revealed a significant main effect of spatial frequency, $F(1,31) = 34.82$, $p < .001$, $\eta_p^2 = .53$. Accuracy for low frequency videos ($M = 0.96$, $SD = 0.60$) was significantly reduced compared to full frequency videos ($M = 1.79$, $SD = 0.56$). The main effect of change type found in the previous experiment was not replicated, $F(1,31) = 0.10$, $p = .75$. However, most importantly, the analysis revealed the predicted change type by spatial frequency interaction, $F(1,31) = 18.42$, $p < .001$, $\eta_p^2 = .37$

Follow-up planned comparisons indicated that, as predicted, featural changes were significantly more difficult to detect when videos were low-pass filtered compared to when full frequency information was present, $t(34) = 9.11$, $p < .001$, $d = 1.54$. Somewhat unexpectedly, configural changes were also significantly affected by low-pass filtering, $t(34) = 2.60$, $p = .01$, $d = 0.44$. However, as the significant interaction and the effect sizes indicate, the effect of high spatial frequency removal was much larger for featural changes than for configural changes.

As in Experiment 1, participants displayed heightened sensitivity to featural information over configural information despite the fact that configural changes were, objectively speaking, significantly greater than featural changes. In the full frequency format, participants were significantly more sensitive to featural changes ($M = 1.97$, $SD = 0.69$) than configural changes ($M = 1.60$, $SD = 0.69$), $t(34) = 3.19$, $p = .003$, $d = 0.54$.

For the most part, the results of Experiment 2 were consistent with our predictions: participants' ability to process featural information in action, and thus their ability to detect changes in local action detail, is dramatically reduced when high spatial frequencies are removed from the stimulus. Although people selectively attend to featural information over configural information, when videos were low-pass filtered, our participants were significantly less skilled at systematically distinguishing featural changes. High spatial frequencies are thus crucial for the processing of featural information residing in details within the motion stream. While this finding may seem intuitive to the point of appearing trivial, it is important to consider these results as part of the broader package of results across both experiments. We have identified one source of information in human action that is selectively attended to even when it involves relatively minor changes to the motion stream represented in high spatial frequencies – featural information – and another source that is not attended to even when it involves relatively large changes to the motion stream represented mainly in low spatial frequencies, and is also sensitive to orientation – configural information.

Interestingly, the processing of configural spatial trajectory information in our videos was also impaired when high spatial frequencies were removed. Importantly, however, this effect was significantly smaller than the effect for featural changes. There are at least two possible explanations as to why sensitivity to configural changes in our videos would be reduced by low-pass filtering: (1) their detection does rely somewhat on high spatial frequency information, and (2) our low-pass filter may have been too extreme. Regarding the first possibility, it is entirely plausible that some high frequency information is used to detect configural changes: we could imagine a possible strategy whereby noticing, for example, the detailed shape of the elbow implies the arm must be following a specific trajectory through space. Regarding the second possibility, the radius of the Gaussian filter we used in order to remove high spatial frequencies may simply have been set so wide that we removed some relatively low spatial frequency information in the process. Indeed, some of our head trajectory changes would have registered in higher spatial frequencies than our arm or torso trajectory changes, and thus their detection would have been difficult if a wide range of high frequency information was removed. In either, case, it is important to remember that in terms of spatial frequency, these two types of information exist on a continuum. What is critical is the relative *magnitude* of the effect for featural and configural change detection. The Experiment 2 findings clearly indicate that detection of featural information in human action is *especially* reliant on high spatial frequencies compared to configural information.

Overall, these findings help to confirm the meaningfulness of the featural versus configural distinction in the action domain. These two sources of information are not only differentially attended to and differentially sensitive to orientation, but their processing also relies on relatively distinct spatial frequencies. In addition, these findings also showcase similarity between action processing and face processing: in both domains, key featural detail relevant to discrimination can be decoupled from configural information to which observers are also sensitive in making discrimination judgments.

6. General discussion

Although rapid and accurate identification of action is crucial for everyday social and cognitive functioning, relatively little is known about processes that serve action identification and discrimination. Any action can be analyzed along a computationally immense number of dimensions, and it is unknown which dimensions people actually utilize when processing human action. Drawing analogies from face-processing research devoted to answering similar questions, we examined the possibility that observers may key on highly relevant detail – featural information – as well as more relational spatial trajectory information – configural information – for purposes of action identification and discrimination. We found that observers are indeed sensitive to both sources of information for these purposes. At the same time, our findings provide clear evidence that featural action information is elevated in adults' processing relative to configural action information. People more readily detect changes to featural detail across the board – regardless of whether action is viewed upright or inverted – and this is so even when configural changes are larger in terms of actual objective, physical change. We also found that processing of configural information is especially reliant on processing action in the upright format. Featural information, on the other hand, is especially reliant on high spatial frequencies for accurate processing. Our findings across these two experiments are consistent with how featural and configural 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 different domains. These results are the first of their kind highlighting the sources of information people readily make use of when discriminating displays of everyday, dynamic human action.

6.1. Selective attention to featural information in action

In our experiments, changes in featural detail within action impacted participants' detection of action differences to a markedly greater degree than changes in configural spatial 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 featural information, as our configural changes were objectively larger changes in physical magnitude. Thus, although the two changes types differed noticeably in one direction objectively, they differed noticeably in the opposite direction *psychologically*. In sum, adults are indeed sensitive to both featural and configural action information, but selectively attend to featural information.

Although rarely discussed explicitly, careful examination of findings in the face processing literature reveals a selective attention effect for featural information that parallels what we have seen in the action domain. In a notable recent exception to the neglect of this important issue, Yovel and Kanwisher (2004) directly controlled for this selective attention effect in their experiment by matching featural change faces and configural change faces in upright discriminability. However, in order to match these two change types on discriminability, they selected configural change faces that were significantly greater than featural change faces on an objective measure of the amount of physical change (Euclidean distance), underscoring adults' selective attention to featural information in the face. Because featural information is elevated in processing, equating featural changes and configural changes requires exaggerating configural changes and making featural changes very subtle.

Recent eye tracking research investigating gaze changes during action viewing meshes nicely with the selective attention to featural detail we consistently observed. 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 ballistic arm trajectories. For example, while watching a person grasp a can of soda and put it in the fridge, eye gaze shifts from the hand to the can of soda before the hand makes contact with the can. Similarly, as the can is subsequently moved to the shelf, gaze shifts from the can to the shelf prior to the can's arrival on the shelf. In our stimuli, featural 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 similarly help to explain their relative difficulty detecting configural changes. The spatial changes in our stimuli tended to occur outside of the hand-object contact region, and thus would be challenging to detect if adults' attention was guided by automatic tracking of predictive hand/object relations. However, trajectory information must be processed in order to appropriately predict the actor's goal and shift gaze, so this information is still utilized, and our experiments indicate that changes in this type of information are discriminable.

It is also important to relate our findings to other features of event perception that have recently been identified. Newtson and Engquist (1976) provided evidence that information contained at segment boundaries is remembered better than information contained at non-boundaries. Recent evidence also suggests that when people segment text narratives of action events, there is a higher probability that they will identify a segment boundary if an object is contacted than if no object contact occurs (Speer, Zacks, & Reynolds, 2007). In our stimuli, featural changes occurred surrounding points of object contact, while configural changes tended to occur independent of object contact. Perhaps participants were more sensitive to changes in featural information because they occurred closer in time to object contact, and thus closer in time to a segment boundary, which may receive more attention. However, it remains an empirical question as to whether or not object contact actually does cue segment boundaries in people's processing of live human action. Furthermore, it is not yet established in the literature as to whether segment boundaries are best characterized as single temporal points in the motion stream or as regions of space-time in the motion stream. If the former characterization is true, then the selective attention effect we observed for featural action information may be an artifact of online action segmentation. If the latter characterization is true, however, then configural action information can be identified within the boundary region, and the featural versus configural distinction is an important distinction over and above online action segmentation. Future studies of both phenomena will help address this exciting issue.

All in all, eye-tracking research, action segmentation research, and our own findings provide converging evidence for selective attention to featural detail in human action processing. Moreover, all point to the region of hand/object contact as a particularly rich source of information for action and/or goal identification.

6.2. Clarifying the featural/configural distinction in action

Our contention is that processing of featural and configural information in action depends upon distinct stimulus properties. Processing of featural information relies heavily on the identification of local detail registered in high spatial frequencies. When high spatial frequencies are removed, as they were in Experiment 2, people have a difficult time systematically discriminating actions based on this source of information. People do not have as much trouble processing configural action information without high spatial frequencies, as configural information tends to register in lower spatial frequencies. A different manipulation highlights a processing characteristic that is more unique to configural information: inversion. Configural information appears to trade more heavily on a prototypical, upright action template. Our conceptualization of this template is similar to the template proposed to exist for the human body (Reed et al., 2003), but differs in that it also includes a dynamic component. Although previous research has demonstrated that dynamic relational processing is disrupted when action is inverted (e.g., Shipley, 2003), the current research further clarifies that a large source of this difficulty is due to the fact that configural information is processed more effectively when actions occur in their typical orientation. When action is viewed upside-down, processing of configural spatial trajectory information is significantly impaired. In contrast, processing featural action information is relatively insensitive to orientation, as it appears to be processed in a piecemeal fashion.

It is important to point out that low-pass filtering disrupted the uptake of both featural and configural sources of action information, but the effect was unevenly distributed across the two sources. Note that such effects are not unique to action: inversion at times affects processing of featural information in the face as well as configural information (e.g., Leder & Carbon, 2006; Yovel & Kanwisher, 2004). Importantly, these effects highlight that the distinction to be made between these two sources of information is one of *degree* more than of *kind*. In any domain, there is intrinsic overlap between featural and configural information, as detailed elements can always be re-described in terms of their spatial properties, and configural information can be re-described in terms of details. In the face, for example, the nose can be described in terms of spatial properties (e.g., the width of the nostrils in relation to the height of the nose bridge). Similarly, non-canonical parts of the face can at times be described as detailed elements, as often occurs in medicine (e.g., the area between the nose and the mouth is known as the 'philtrum'). Ultimately it is best to conceptualize these two sources of information as lying on a continuum, particularly in terms of spatial frequency. These results tell us that featural action information is especially reliant on relatively high spatial frequencies compared to configural action information.

6.3. Action and the face: both examples of a domain-general dual processing approach?

Our findings lend support to the idea that a featural/configural 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; Reed et al., 2003). When we watch other people's action, at times we must be quite sensitive to subtle properties of their 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 simultaneously in parallel, could be an efficacious solution. In partial support of such a domain generality view, evidence suggests the same or similar brain regions implicated in processing featural and configural information in the face are also involved in expert object identification (cars, birds, and novel "Greebles", Gauthier, Skudlarski, Gore, & Anderson (2000); Rossion, Gauthier, Goffaux, Tarr, & Crommelinck (2002)). Although not critical to supporting this domain generality view, it would be interesting to explore whether brain regions implicated in the featural/configural distinction in face processing and expert object identification are also implicated in processing these distinct sources in human action.

7. Future investigations

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

Another question for the future concerns how entrenched the selective advantage for processing featural detail tends to be. It is potentially of interest whether it can be readily shifted via simple manipulation such as explicit instruction to attend to spatial and timing changes or contextual manipulations that shift an observer's attentional set. Selective attention to featural information over configural information may be a flexible difference, and may influenced by the particular context and task demands (Morrison & Schyns, 2001).

The developmental origins of adults' action processing profile concerning featural versus configural information is also ripe for investigation. For face processing, a small amount of evidence indicates that processing of configural information develops more slowly than processing of featural information. For instance, Mondloch et al. (2002) found that children's accuracy for detecting featural changes in the face improved between 6–10 years of age, and that 10-year-olds' accuracy was equivalent to that of adults. Accuracy for configural changes developed more slowly, and children's accuracy for detecting such changes was still significantly poorer than adults' at 10 years of age. Whether a similar developmental pattern holds for human action is of significant interest. Along these lines, recent evidence suggests that 12-months-old infants show gaze shifts in action observation strikingly similar to those found with adults (Flack-Ytter, Gredebäck, & von Hofsten, 2006). As mentioned earlier, such predictive gaze shifts may focus attention to featural detail and away from configural spatial properties, and thus by 12 months selective attention to featural over configural information may already be present. Moreover, infants as young as 9 months are known to be sensitive to local details of hand/object contact (e.g., grasping versus back-of hand-contact) in construing action as goal-directed (e.g., Woodward, 1999). Whether infants' sensitivity to relevant featural detail within action extends to other contexts is yet another interesting question for future investigation.

8. Conclusion

These experiments showcase people's skill at flexibly homing in on highly relevant details within motion to assist in identifying the nature of others' activity. At the same time, however, our findings confirm that adults systematically track configural characteristics of motion for identification and discrimination purposes. Psychologically speaking, changes in action-relevant featural detail loom large in adults' processing, whereas changes in configural characteristics of the same physical magnitude are less detectable. Moreover, processing of featural versus configural action information can be impacted in contrasting ways by manipulations such as inversion versus low-pass filtering, confirming that this is a meaningful distinction in adult action processing. The pattern of findings that has emerged in these experiments directly parallels those documented in adults' processing of faces. This points to the possibility that the featural/configural dual-source processing approach observed in both action and face processing arises through domain-general mechanisms subserving the acquisition of observational expertise.

Acknowledgments

This project was supported by a grant from the National Science Foundation under Grant No. BCS-0214484. We would like to thank Bridgette Hard, Ulrich Mayr, Bertram Malle, Meredith Meyer, Bridget Klest, and Emily Neuhaus for insightful and helpful comments and discussions. We would also like to thank Amanda Altig, Stephen Boyd, Lindsey Henson, and Sarah Wilson for invaluable help with data collection. Finally, many thanks to all of the participants who helped make this research possible.

References

Baird, J. A., Baldwin, D. A., & Malle, B. F. (2008). Parsing the behavior stream: Evidence for the psychological primacy of intention boundaries. Submitted for publication.

Baldwin, D. (2005). Discerning intentions: Characterizing the cognitive system at play. In B. D. Homer & C. S. Tamis-LeMonda (Eds.), *The development of social cognition and communication* (pp. 117–144). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Baldwin, D. A., & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences, 5*, 171–178.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development, 72*, 708–717.

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25*, 281–316.

Bertenthal, B. I., Proffitt, D. R., & Cutting, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motions. *Journal of Experimental Child Psychology, 37*, 213–230.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Collishaw, S. M., & Hole, G. J. (2002). Is there a linear or a nonlinear relationship between rotation and configural processing of faces? *Perception, 31*, 287–296.

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

Flack-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience, 9*, 878–879.

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., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*, 191–197.

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.

Golinkoff, R. M., Chung, H. L., Hirsh-Pasek, K., Liu, J., Bertenthal, B. I., Brand, R., et al (2002). Young children can extend motion verbs to point-light displays. *Developmental Psychology, 38*, 604–614.

Hard, B. M., Lozano, S. C., & Tversky, B. (2006). Hierarchical encoding of behavior: Translating perception into action. *Journal of Experimental Psychology: General, 135*, 588–608.

Hillger, L. A., & Koenig, O. (1991). Separable mechanisms in face processing: Evidence from hemispheric specialization. *Journal of Cognitive Neuroscience, 3*, 42–58.

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception, 31*, 1221–1240.

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

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2003). Expert face processing requires visual input to the right hemisphere during infancy. *Nature Neuroscience, 6*, 1108–1112.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 2*, 513–536.

Leder, H., & Carbon, C.-C. (2006). Face-specific configural processing of relational information. *British Journal of Psychology, 97*, 19–29.

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

Meltzoff, A. N. (2002). Imitation as a mechanism of social cognition: Origins of empathy, theory of mind, and the representation of action. In U. Goswami (Ed.), *Handbook of childhood cognitive development* (pp. 6–25). Oxford: Blackwell Publishers.

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

Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin and Review, 8*, 454–469.

Murray, J. E., Yong, E., & Rhodes, G. (2000). Revisiting the perception of upside-down faces. *Psychological Science, 11*, 492–496.

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

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology, 12*, 436–450.

Pavolva, M., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception and Psychophysics, 62*, 889–899.

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

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169–192.

Rossion, B., Dricot, L., Devolder, A., Bodart, J. M., Crommelinck, M., de Gelder, B., et al (2000). Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 12*, 793–802.

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.

Saylor, M. M., Baldwin, D. A., Baird, J. A., & LaBounty, J. (2007). Infants' online segmentation of dynamic human action. *Journal of Cognition and Development, 8*, 113–128.

Sergent, J. (1986). Microgenesis in face perception. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (pp. 17–73). Dordrecht, The Netherlands: Nijhoff.

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

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science, 18*, 449–455.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471–491.

Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior & Development, 22*, 145–160.

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

Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron, 44*, 889–898.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., et al (2001). Human brain activity timelocked to perceptual event boundaries. *Nature Neuroscience, 4*, 651–655.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General, 130*, 29–58.