Biological Motion: An Exercise in Bottom–Up vs. Top–Down Processing

Basileios Kroustallis

University of Crete

Biological motion is the phenomenon of recognizing a human form out of moving point-light dots, where both bottom–up and top–down processing mechanisms have been reported. This study reviews available psychological and neuroscientific evidence, and it assesses attempts either to assimilate biological motion to other structure-from-motion cases (bottom–up) or to include biological motion into a visual "social cognition" subsystem (top–down). While neither theoretical option seems to accommodate all relevant psychological results, the study proposes that biological motion may be an object recognition task, inside the framework of Pylyshyn's (1999a) sequence of data-driven and cognitive mechanisms. This implies that a bottom–up object construction out of two-dimensional stimulus information precedes a top–down, but emotionally significant categorization of a particular human movement. Recognition of biological motion may be an example of visual processing in general.

Key words: biological motion, vision, object recognition

In recent years, neuroscientific findings and psychological experiments on vision have transformed the existing knowledge about that sense. Visual processes have been studied meticulously and in depth, although there are still problems left unresolved and disputed. The issue arises whether vision is independent of general cognition or whether external influences (expectations, familiar perceptions) intervene and corroborate the stimulus information. In other words, it is questioned whether visual perception is an autonomous component of cognitive processing (Fodor, 1983) or whether cognitive expectations emerge and influence the visual output (Gregory, 1966; Rock, 1983).

The complicated tasks that vision accomplishes cannot be only the result of retinal input. Visual perception may be bottom–up (or data-driven), but helped by means of natural constraints, which are assumptions about the external world that limit possible outcomes of a visual process according to what is frequently experienced. For example, the spatial continuity assumption denotes that markings on a surface are often spatially organized, arranged in curves, lines or possibly more complex patterns (Marr, 1982, p. 49). The construction of a three-dimensional structure out of a two-dimensional image seems to rely upon these general assumptions. These constraints are thought somehow to be built-in into the visual system, and agents cannot choose to employ them or not, unlike with beliefs or expectations.

Nevertheless, a function as complex as vision is expected to involve influences stemming from general knowledge or agent expectations. In so many cases agents seem to see what they want to see, and even in ordinary object recognition, there seems to be a later stage, where visual information is compared to memory-stored object types, and object identification follows (cf. Biederman, 1993).

This study will examine the phenomenon of biological motion, a case where human recognition is achieved by means of motion information, in the absence of form perception. In the following two sections it will be shown that biological motion may equally support both bottom–up and top–down approaches. Neuroscientific data and psychological results can be interpreted either as instances of shape recognition through motion extraction (bottom–up) or as part of a visual subsystem concerned with identification of "social" characteristics, such as facial expression and purposeful action (top–down). In the final section, alternative ways will be presented to integrate that evidence: it is proposed that a current object recognition scheme (Pylyshyn, 1999a), which endorses a sequence of early, data-driven processing followed by cognitive penetration, can best interpret the otherwise ambivalent findings on biological motion. This case does not provide a final judgment on the matter of bottom–up vs. top–down involvement, but can offer an insight on the various information-processing mechanisms in vision.

Biological Motion as Bottom–Up Processing: The Structure-from-Motion Case

An experimenter attaches lights to an actor's body joints (ankles, knees, hips, shoulders, elbows and wrists). The actor is filmed in the dark, and the film is shown to different subjects, who are asked to report what they see. When the frame is still, the subjects perceive the lighted points in the joints, but they only appear to them as meaningless sets of dots, like stars in the sky.

However, when the actor moves, they can suddenly recognize a person moving and her specific movement type (walking, jumping, etc.), reporting a vivid perception of that fact. This phenomenon of human recognition through motion without adequate form perception is termed biological motion (Johansson, 1973). Subjects recognize not only moving humans, but also dancing couples or the gender of a walker (Cutting and Proffitt, 1981), while infants at the age of four months attend biological motion patterns for a longer duration than random patterns of moving lights (Fox and McDaniel, 1982).

Biological motion seems to be a case of apparent motion — an impression of motion without object motion. Successive film frames give the impression of a moving object on the screen, just like a frame sequence in a film. In general, apparent motion is divided into two distinct types, according to the time necessary for its recognition: short-range and long-range motion (Braddick, 1974). Small displacements of random-dot patterns in successive frames are perceived as object motion, and reveal a short-range motion mechanism, which is deactivated (and motion perception stops) if the frame alternation rate is longer than 80 ms.

Classical cases of apparent motion, on the other hand, utilize large individual objects as stimuli, and significant frame displacements. The interval between successive frames, as well as the frame duration, has also to be of significant duration (circa 500 ms and 300 ms respectively) for motion to be recognized. So, these are appropriately considered a result of the long-range motion mechanism. Motion in the latter case is also perceived even when the different frames are presented to different eyes, and since this involves the integration of information from the two eyes, it seems that the long-range motion mechanism occurs rather late in visual processing. The short-range system does not respond to that information, and it is considered to involve early processing.

Biological motion was initially thought to be a high-level process, because of the inherent complexity of the percept, so that recognition of the moving form could not be explained by early stimulus information alone. However, subjects recognize the displays shown in an interval less than 60 ms between successive frame presentations (Johansson, 1973). Furthermore, this percept is robust, since changes in dot contrast polarity or spatial frequency do not alter motion perception (Ahlström, Blake, and Ahlström, 1997).

To prove that biological motion involves essentially low-level processing, Mather, Radford, and West (1992) tested their subjects for motion recognition using frame displacement intervals between 60 ms and 100 ms. They argued that if biological motion were a high-level process, there would be no difference in subject performance, since the interval threshold for long-range motion recognition is much higher, up to 500 ms. However, subjects' perfor-

mance deteriorated between 60 ms and 100 ms intervals and, therefore, it was concluded that biological motion is a low-level process.

This claim has been questioned. Thornton, Pinto, and Shiffrar (1998) showed that subjects were able to perceive biological motion with approximately 75–80% correctness for an interval larger than 100 ms, something that denotes long-range motion recognition. On the other hand, Chatterjee, Freud, and Shiffrar (1996) examined the individual frame duration factor, which had not been considered in the Mather et al. (1992) study. They found that longer frame duration leads to better performance in biological motion, a result that could be interpreted in favor of a high-level processing, since longer frame duration may permit cognitive factors to intervene in the response.

So, if biological motion is a case of long-range motion processing, there must be an explanation of the production of the phenomenon. It is usually assumed that the distinction between short and long-range motion is the equivalent of bottom–up and top–down processing respectively. Short-range mechanism can be characterized as low-level processing. However, high-level processing may not only be associated with top–down influences. Biological motion may belong to the long-range motion system and, at the same time, be a construction out of stimulus characteristics.

This is the case for "structure-from-motion" phenomena, shape recognition of geometrical structures by means of motion information. Biological motion may be initially assimilated to this process and placed in a broader object recognition framework. There are studies that link the two cases: for example, under dim-light conditions, perception of coherent motion is unaffected, but perception of biological motion and structure-from-motion is equally and severely impaired (Grossman and Blake, 1999).

Computational studies in motion perception solve the problem of structure-from-motion by means of an application of natural constraints into two-dimensional image information, just like shape formation, results from shading, depth and texture. This is essentially a bottom–up process from specific information to a unified percept. Subjects prefer to perceive rigid, object-like structures in the environment. Even though immediate stimulus information may consist of diverse elements, it is assumed that the observers attempt, whenever possible, to interpret moving things as rigid structures, under the least available evidence of four corresponding object points in different views of the moving structures. Consequently, a natural constraint for perception of objects in motion is the rigidity assumption (Ullman, 1979).

In biological motion, unlike other structure-from-motion cases, there is not a single object direction, since the different limbs (hands, feet) can seem to move independently in different directions (back or forth). Nevertheless, it is proposed (Webb and Aggarwal, 1982) that two points are sufficient to

provide the perception of a rigid part, and the constraint is appropriately modified to reflect not global, but local rigidity (motion of separate limbs). Johansson (1973) proposes that this modified assumption is satisfied in biological motion by means of perceptual vector analysis: subjects purportedly analyze complex motion of the specific body parts into components, and then structure them in order to form a single percept. Specifically, the motion of the hips and the knees is compared and a common direction is kept, a horizontal (translatory) motion, while the remaining component (pendulum motion) is discarded. The same common direction (translatory motion) is found in all pairs (knees–ankles, and shoulders–elbows, elbows–wrists, in that order). This hierarchy of relative motion comparison gives global horizontal motion perception.

Webb and Aggarwal (1982) advanced an alternative motion perception solution based on a fixed axis assumption. All movement occurs about an axis, which is fixed in direction for short periods of time. Local, independent sets of movement are constituted around that axis, comprised of only two visible points (which represent a rigid part). Joints act as a common point between two interrelated parts, and they can unify an overall structure. Another proposal places constraints not in the spatial, but in the temporal characteristics of separate limb movement, e.g., the relative frequency comparison of separate motion-cycles (Bertenthal and Pinto, 1993).

All the above models view biological motion as a modified case of structure-from-motion. However, biological motion, unlike structure-from-motion cases, is not susceptible to the aperture problem, the integration of local motion. Subjects viewed figure renditions of different objects, such as car, scissors and walkers, through apertures. They were able to identify walking humans with hidden joints through these apertures, but failed to identify inanimate objects under similar conditions (Shiffrar, Lichtey, and Chatterjee, 1997).

At the same time, converging neurophysiological results lead to the same conclusion. The area V5 (or MT), which is responsible for object motion, seems to be also the center of structure-from-motion perception (Andersen and Bradley, 1998), but this same area responds indifferently to biological motion patterns (Grossman, Donnelly, Price, Pickens, Morgan, Neighbor, and Blake, 2000). On the other hand, fMRI indicates that biological motion patterns activate different brain areas than presented displays of inanimate things (Grèzes, Fonlupt, Bertenthal, Delon–Martin, Segebarth, and Decety, 2001). Under these circumstances, it is difficult to equate the two processes. Long-range motion recognition, rather than involving complex form reconstruction, may employ either an immediate grasp of complex perceptual information (Cutting and Proffitt, 1981) or the exertion of various cognitive influences, such as spatial attention or memory.

The second alternative is invoked by the phenomenon of inverted biological motion. While the subjects can recognize a moving display of a human form, they have difficulty recognizing it when it is positioned upside-down (Sumi, 1984). This is a sign that bottom–up processing is not the only factor operating in biological motion, for if that were the case, one could use the local principles for motion of body parts and still recognize the display.

There is additional evidence supporting this proposal. Sensitivity to biological motion increases rapidly with the number of illuminated points in the display, far more rapidly than recognition of simple motion. This recognition seems to be influenced by the nature of the stimulus (Neri, Morrone, and Burr, 1998). Masking studies lead to similar conclusions. In a set of psychological experiments, random dots were superimposed on an original pattern, so that the spatial relation between the lights on the joints was lost (and no local rigidity constraint could operate). Nevertheless, the subjects could still perceive the biological motion pattern that was evident without the additional dots (Ahlström et al., 1997; Thornton et al., 1998).

Even computer theorists, who advocate a local, bottom–up processing by means of natural constraints, take care not to exclude a role for previous knowledge. Johansson (1973, p. 210) asserts that experience does not *necessarily* determine biological motion recognition, though he admits that the percentage of almost 100% correct answers in his experiments shows that in most cases experience plays a role. Webb and Aggarwal (1982) accept that higher-level knowledge may aid interpretation of human movements. Similarly, although Mather et al. (1992) claim that their low-level explanation is sufficient for biological motion recognition, they admit it may not be necessary.

Biological Motion as Top–Down Processing: The Social Cognition Case

Ungerleider and Mishkin (1982), based on a wide range of neurophysiological findings, proposed a distinction between two streams of visual processing from retina to the visual cortex and higher cortical areas in the brain, each responsible for the separate processing of form perception and motion recognition. Milner and Goodale (1995), reviewing specific cases of neuropsychological deficits, argued that even when the brain cannot use information for object recognition, it can nevertheless process visual object information that facilitates action. An example of the latter was that of blind patients unable either to recognize or to point to a stimulus but with the ability nevertheless to post a letter into an unseen slot. Similar cases led those researchers to a modification of the two pathways hypothesis, which was formulated as a distinction between a network responsible for perceptual recognition, and a "recognition for action" stream respectively.

If biological motion belonged to a motion pathway, and it was a case of object detection through motion recognition, there would be corresponding neural areas activated during perception of biological motion displays. Nevertheless, neuroscientific data show that although object detection may be determined by motion recognition, motion recognition may not be necessary for biological motion. Clinical experiments with human subjects show that there is a difference between ordinary motion and biological motion. Patients with impairment in motion mechanisms (and are thus unable to perceive motion) can recognize biological motion patterns (Vaina, LeMay, Bienfang, Choi, and Nakayama, 1990). Conversely, subjects can be quite competent in ordinary motion perception without recognizing a form out of the biological motion display. These subjects perceive that not all dots in a pattern move in the same direction, but they cannot reconstruct the shape of the pattern. This has been called "visual motion agnosia" (Cowey and Vaina, 2000), where subjects see the motion but not what it represents.[1]

The above neuropsychological findings are important, but they could be characterized as inconclusive, since it is not known whether the observed deficit is a direct consequence of a malfunctioning area or is an outcome of the connecting role that this brain area has in a neural network comprised of many areas. Nevertheless, electrophysiological studies in macaque monkeys (e.g., Oram and Perrett, 1994) indicate a distinct brain area activated during biological motion displays. That area consists primarily of cells in the posterior end of the superior temporal sulcus (STS), an area superior and anterior to V5 (the area of visual movement recognition). Neuroimaging studies (Grèzes et al., 2001; Vaina, Solomon, Chowdhury, Sinha, and Belliveau, 2001) confirm the role of STS in human perception of biological motion.

However, STS has a special place in the different neural pathways of object recognition processes. The activation of that area in terms of its input and its function cannot be attributed exclusively either to the motion or the form pathway, but is rather a synthesis of both (Vaina et al., 2001).

On the other hand, even if STS is the area that integrates different object features into a coherent percept, this does not solve the problem of the corresponding synthesis of the different processes in the functional level. Biological motion cannot be assimilated to structure-from-motion phenomena, and it is not ordinary form recognition.

Another alternative emphasizes that the categorization processes that intervene in the perception of a human form are not always the result of an object detection model. Vision may contain other subsystems apart from object recognition, some of which may also reflect cognitive influences.

---

[1]The term being borrowed from prosopagnosia, the inability to recognize faces, while still recognizing individual characteristics.

Specifically, in the field of face recognition, the top–down mechanism is expressed by context identification (i.e., subjects are influenced by the environment where that face is viewed), which compensates for lack of specific, face feature extraction (Palmer, 1975). Face recognition is hampered when the faces are presented inverted, just like perception of inverted displays in biological motion (Bartlett and Searcy, 1993). Furthermore, biological motion recognition seems to support an activation of the fusiform gyrus, an area associated with face processing information (Grossman and Blake, 2002).

Nevertheless, the two phenomena are not the same. Superior temporal sulcus responds only to changeable face characteristics, such as a gaze or an expression, and it fails to respond to a static face as a whole, as other areas do (Haxby, Hoffman, and Gobbini, 2000). On the other hand, prosopagnosia is compatible with unimpaired biological motion perception (Vaina et al., 2001).

Another visual subsystem that employs cognitive influences is "social cognition," the interpretation of visual signals either as intended communication or intentional behavior towards a certain object (purposeful action). Perception of gaze direction is a communication sign, while the staring gaze of one animal to another can be interpreted as a possible threat. Hand movements towards objects denote goal-directed actions, and bodily movements without evident form perception can be regarded as a tendency to act in a specific way (Allison, Puce, and McCarthy, 2000).

Clinical studies reveal that social cognition is dissociated from ordinary motion perception and object recognition. Autistic children interact very poorly with their immediate, social environment. Nevertheless, they are able to detect motion, since they can perceive the gaze direction of another person, although they cannot use this to infer the other's mental state. They lack the connection with the object towards which the gaze is directed (Baron–Cohen, Campbell, Karmiloff–Smith, Grant, and Walker, 1995). On the other hand, in a disorder called Williams syndrome, although patients have impaired motion abilities, they can succeed in social cognition tasks. For example, they can interpret the mental state of a person when they are presented with a photograph of his eyes (see Adolphs, 1999). These subjects also perform well with biological motion displays.

Adolphs (1999) has proposed a distinct network of "social" brain areas to account for these findings, in which STS plays an important part. And neurophysiological studies verify the results of pathological cases, and the same characteristics that define social cognition (interpretation of gaze direction, mouth movements, goal-directed hand movements) also activate STS cells (Allison et al., 2000). Since STS also responds to biological motion patterns, biological motion is claimed to be essentially a case of social cognition.

However, it is not certain that biological motion fits perfectly into this network. There are some details that don't match this scheme. While STS responses may serve many psychological functions (including social cognition issues), this does not necessarily imply that all those separate functions help to explain the biological motion case. On the other hand, not only STS is activated during viewing of biological motion displays. Both the inferior temporal cortex and the intraparietal cortex are involved (Grèzes et al., 2001), which have different visual functions.

Therefore, identification of the function of biological motion with social cognition is not adequately supported. Psychological studies also point against a "social" interpretation of the biological motion case. Experiments with cats (Blake, 1993) showed that they can also perceive biological motion patterns — of other cats. This occurred in spite of the fact that cats hunt at night and are not very social animals, which shows that communicatory purposes do not seem to be necessary for biological motion recognition — although alternative proposals for that ability, such as the need to recognize a potential predator or prey, have been proposed (see Wallisch, 2001).

Furthermore, even when subjects have to recognize different activities performed by agents in biological motion displays, they do not show a selective preference towards "social" activities. Dittrich (1993) associated top–down processing with the recognition of specific human action types, locomotory (walking or jumping), instrumental (task execution with the help of an instrument), and social actions (dancing, greeting). In those experiments, a selective response of the subjects was recorded with respect to the three categories. The recognition rate was higher in the first type (locomotory movements), followed by responses concerning the "social interaction" category, and lower ratings were found in instrumental movement recognition. If biological motion could be essentially explained as social cognition, a corresponding preference would be attested in subject performance. But the most common and less "social" kind of actions (locomotory actions) achieved a higher recognition rate. Therefore, category recognition does not work as an all-purpose, top–down biological motion influence. It is more a result of a particular action and its possible significance for the subject.

Object Recognition and Biological Motion

The fact that biological motion can be neither explained as a bottom–up structure-from-motion case nor as part of a top–down social cognition network may lead to the view that biological motion is *sui generis*, irreducible to any known visual case. Nevertheless, the apparent conflict does not take into account that an object recognition system may also involve top–down processes, where the necessary (but not easily fulfilled) requirement is the

exact connection with stimulus-driven processes. Both bottom–up processing and top–down influences may be integrated inside the framework of an object recognition task. There are (at least) two possible ways for this to occur, and they will be examined in turn.

Dittrich's (1999) proposal was that such integration occurs within an "interactive encoding model" for biological motion. A parallel working memory circuit creates the impression of a human action type, which is attested in category recognition. This memory circuit interacts with bottom–up processes, which produce a three dimensional movement out of two dimensional stimulus information. However, top–down motion classification is not a later product, which is somehow added to an already-formed percept. A two-part process, with specialized "biological motion detectors," which process form out of motion and later categorize the percept as a meaningful object, is made unnecessary by the low frame interval threshold. Agents recognize biological motion displays, so there is no need for elaborate bottom–up processes. On the contrary, motion classification may occur at an early stage, where visual motion cues (e.g., direction, speed) are combined with semantic characteristics (e.g., greeting). This interaction of an object detection system with category classification through memory is argued to be responsible for biological motion perception.

Another argument that Dittrich (1999) brings in favor of his model is that biological motion responses may be assimilated to other psychological studies (Schyns and Oliva, 1994) on general scene identification, where a photograph in a masked presentation can be identified as a particular scene type (e.g., a bridge or a music hall) within a duration as short as 45–135 ms. This identification is independent of information concerning specific object features, so it is considered to be a result of top–down processing. Dittrich (1999) makes a clear parallelism: early, top–down scene information mirrors early action type classification in biological motion.

However, it does not suffice to prove that scene information is as early as movement classification for the parallelism to work. It has to be found that early scene information actually influences later object identification, so that the corresponding claim in biological motion that early action classification intervenes on object reconstruction will be supported by the former fact.

Accordingly, there are two different proposals on object and scene recognition (Henderson and Hollingworth, 1999). Both claim that the semantic coherence of a scene enhances sensitivity to object identification, e.g., it is easier to recognize a briefcase in an office than on a highway. However, the first proposal asserts that in the first case an already-formed object structural description is matched against long-term memory representations (Kosslyn, 1994; Ullman, 1996). Scene knowledge influences only the criterion used to determine that a particular object type is present, without directly determin-

ing the perceptual analysis of the object token. In contrast, the second view explains that case by claiming that expectations derived from scene-specific information influence directly the perceptual analysis and the formation of an object's structural description (Boyce, Pollatsek, and Rayner, 1989). In both cases, there is said to be an earlier, top–down influence of scene recognition upon the object identification process.

In a series of experiments, Hollingworth and Henderson (1998, 1999), tested this claim. They argued that there is a scene identification influence on the overall processing, but this kind of information is not *prior* to either object identification or to object detection. When later effects of context (expressed as response bias) are eliminated, there is no advantage in detecting a coffee maker in a room rather than in a street. General scene information influences subjects' tendency to respond appropriately to an object identification task, not to their ability to identify an object, which has already taken place independently. This shows that scene recognition is not unanimously endorsed as an early factor determining object identification. Therefore, a model of biological motion, which endorses an early interaction of action classification with object motion and is based on the former claim cannot stand as it is.

Nevertheless, there is an alternative scheme on object recognition compatible with contextual influences, but contrasted with the interactive encoding model and its emphasis on early top–down categorization. Pylyshyn (1999a) divides vision into three stages: focal attention precedes all data-driven processes, by preselecting a set of salient objects to serve as the primary input to the visual system. An early vision mechanism follows, which admits no cognitive influence or "cognitive penetration." This stage employs available stimulus information, and uses natural constraints, culminating in the construction of a three-dimensional object description. The third step involves classification of this description as a certain object type that is partly the result of extravisual factors. The two later stages reflect the distinction between "see an object" and "see an object as something," i.e., identify an object. This is the turning point between integration of visual information into a unified percept (a result of data-driven processes) and decision on the classification of that percept (top–down influence).

Pylyshyn (1999a) provides supporting evidence for this scheme from different sciences, such as neuroscience, psychology of perception and computer vision. However, there are also contrasting voices. Barsalou (1999), for example, claims that early vision is not a separate stage from general cognition, but only when bottom–up information conflicts with top–down influences (as it may be in phenomena such as the Müller–Lyer illusion), then will bottom–up information dominate the resulting percept.

Therefore, instead of attempting to evaluate this scheme for visual phenomena in general, it is useful to see whether this can be applied to the bio-

logical motion case. A first objection would be that focal attention, which Pylyshyn (1999a) recognizes as an initial top–down factor, might already be influenced by movement categorization for different displays. Schyns (1999) argues that attending to the visual cues that solve a categorization task can change the actual perception of the stimulus. If that is the case, prior categorization by means of memory may allocate attention exactly to a specific percept type, and there would be no reason for later cognitive intervention.

Attention allocation in biological motion refers to the act of choosing between a motion pattern and a completely different pattern. Subjects tested in biological motion recognition were asked to respond simultaneously to a completely different attention-demanding task (Thornton, Rensink, and Shiffrar, 2002). This was meant to distract their attention, and measure the corresponding effect on biological motion recognition. Subjects' response to motion displays was influenced by the allocation of attention to the second task, and it was also observed that performance in the first task (biological motion) was significantly worse in motion patterns characterized as involving top–down processing than in patterns considered an outcome of bottom–up processing. Sensitivity to attention withdrawal seemed to coincide with application of cognitive influences but at the same time, attention allocation was not itself the cognitive factor influencing this performance. There was *additional* information that a certain type of motion pattern involves conceptual-driven influence. In other words, attention seems to accompany existing top–down influences, and it is not by itself a means for action categorization.

A second line of dissent would be to argue that biological motion is a paradigm case of the non-applicability of constraints, such as the rigidity assumption, in this specific case (Dannemiller and Epstein, 1999). However, a single constraint (e.g., rigidity) does not need to apply in every visual case, but there can be different constraint-based explanations. The only requirement for those explanations is that they express restrictions based on visual surface regularities and not on the semantic coherence of a scene, unlike contextual influences (Pylyshyn, 1999b).

These objections concern the applicability of the object recognition scheme to biological motion. However, there is also evidence that biological motion actually involves a sequence of bottom–up and top–down processing. In a recent study (Vaina et al., 2001), subjects were presented with the same pattern of dots twice. In the first display, they were asked to report whether there is an object in the pattern, whereas in the second display they had to denote the overall motion direction of the dots. Performance was different in the two cases, for subjects did not see a moving object in the second display, even though the visual pattern was still the same. The neural areas activated were also different. The results show that subjects responded not to the visual display as such, but to the task which was asked of them. It is as if they

"decided" (under the appropriate instructions) to see the same visual pattern presented either as an object or as only a pattern of moving dots.

It seems that familiar pattern recognition may be one relevant, top–down factor. This is enforced by Dittrich's (1993) movement type results, and it is also shown in a study with children as subjects (Pavlova, Krägeloh–Mann, Sokolov, and Birbaumer, 2001). Researchers found out that their responses are not identical with those of the adult subjects, since children recognize patterns of dogs from motion displays better than human forms, and they argued that this might also happen because dogs are more familiar to children in terms of height and size. This result, in conjunction with cats' responses towards other cats (Blake, 1993), indicate that the essence of the biological motion mechanism may not be the recognition of a human form, but the recognition of a familiar living form.

Although it could be argued that animals might have been endowed (or hardwired) with the ability to recognize their prospective predator or prey, even under diverse circumstances (absence of form perception), the above-mentioned studies indicate that this ability may not be an unalterable built-in assumption. Whereas infants of only four months of age can recognize biological motion patterns, it takes time to reach the peak of their biological motion performance, up to age five, although the respective duration for an ordinary object recognition is not that long. Cats can be trained successfully to respond to human motion displays (Blake, 1993). Biological movement type recognition does not work as another immovable local constraint, but as a top–down influence, modified by experience.

Nevertheless, familiarity by itself cannot account for a phenomenal accompaniment of biological motion. It has already been mentioned that biological motion recognition (unlike structure-from-motion or social cognition cases) is usually followed by a kind of vividness, a vivid impression of a human being walking, jumping and so on (Johansson, 1973; Vaina, 2001). This impression needs to be interpreted, since if this were only the result of the absence of form perception, vividness would be reported in the static figure displays, or in other structure-from-motion cases.

Research results from both neuroimaging and lesion studies show that the amygdala is associated with biological motion recognition. The amygdala is preferentially activated when human body movements are displayed, in contrast with inanimate object motion or random motion. On the other hand, it is not involved in goal-directed action processing, such as hand movements to reach an object, and this counts as evidence against its inclusion in a social cognition network (Bonda, Petrides, Ostry, and Evans, 1996). The general function of the amygdala is acknowledged to be either the production of emotional responses, such as fear or anger, or the recognition of those emotions in facial expressions of other persons (Adolphs, 1999; Allison et

al., 2000). However, since there is no information concerning form perception in biological motion, no facial expression recognition occurs. Therefore, amygdala activation may "reflect the emotional disposition experienced by the perceiver in response to the behavior perceived" (Bonda et al., 1996, p. 9743). This disposition, though, needs to be further defined.

Apart from the previous considerations, additional evidence concerning the amygdala comes from an experiment in which subjects saw a movie with simple geometric shapes in motion, and then were asked to report what they saw. Subjects with lesions in amygdala described the movement in purely geometric terms, whereas subjects with normal amygdala function viewed the scene as a living sequence: "I saw a box, like a room . . . . There was a large triangle chasing little triangle. Finally he went in, got inside the box to go after the circle, and the circle was scared of him . . ." (Heberlein, Adolphs, Tranel, Kemmerer, Anderson, and Damasio, 1998). Therefore, it might be plausibly suggested that the function of amygdala in biological motion is connected with the acknowledgment of a living being in the context of point-light displays, thereby creating a vivid impression of an animate form.

Amygdala activation has also been associated with top–down feedback. Experiments on face recognition showed that recognition of face identity was followed by later face expression recognition, and the latter result was attributed, among other neuronal structures, to an amygdala influence (Sugase, Yamane, Ueno, and Kawano, 1999). It seems that amygdala in general may work as a top–down, modulation factor (Allison et al., 2000). On the other hand, memory influences on classification types can be better assimilated to stored object models which are recalled after object reconstruction (Biederman, 1993) rather than to contextual elements (general scene recognition), simultaneous with object recognition (Dittrich, 1999). That memory influence would also be a top–down processing factor in biological motion recognition.

Biological motion can work as an example of object identification through a sequence of bottom–up and top–down mechanisms. It is neither a structure-from-motion case, which can only be explained by the application of local constraints, nor a generic top–down social cognition case. Biological motion recognition is divided into two parts, the construction of human movement, and recognition of human action. In the first part, the subject's attention is turned towards the biological motion displays, although people do not see a meaningless pattern of dots in motion. A grouping of elements (i.e., the dots) is essential, in order for the subject to perceive a unified moving object. This is accomplished in accordance with the built-in visual constraints, which form an object, while inside that framework different solutions may be advanced towards the problem of extracting a three dimensional movement from two dimensional information. In the second part, recognition occurs

after a decision to classify three-dimensional movement as biological action. This is the vivid impression that subjects report, which can result from a feedback influence from subcortical areas, such as the amygdala, though it also has to include long-term memory areas, where stored human movement types are compared to the actual percept.

The above considerations form a preliminary view of the available evidence, and a methodological effort to accommodate data into a specific scheme. Further work needs to be conducted both to clarify the particular constraints operating in the production of biological motion, and to determine the exact function of the neural areas involved. It has to be noted that even if forthcoming evidence agrees with the above review, this does not imply the validity of Pylyshyn's (1999a) object recognition scheme for all visual phenomena. Nevertheless, biological motion may be an example of how diverse visual information sources can be integrated under a specific object recognition model, and this is a step towards understanding visual processing.

References

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences, 3,* 469–479.

Ahlström, V., Blake, R., and Ahlström, U. (1997). Perception of biological motion. *Perception, 26,* 1539–1548.

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences, 7,* 267–278.

Andersen, R.A., and Bradley, D.C. (1998). Perception of three-dimensional structure from motion. *Trends in Cognitive Sciences, 2,* 222–228.

Baron–Cohen, S., Campbell, R., Karmiloff–Smith, A., Grant, J., and Walker, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology, 13,* 379–398.

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–609.

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

Bertenthal, B.I., and Pinto, J. (1993). Complementary processes in the perception and production of human movements. In L.B. Smith and E. Thelen (Eds.), *A dynamic systems approach to development: Applications* (pp. 209–239). Cambridge, Massachusetts: The MIT Press.

Biederman, I. (1993). Visual object recognition. In A.I. Goldman (Ed.), *Readings in philosophy and cognitive science* (pp. 9–21). Cambridge, Massachusetts: The MIT Press.

Blake, R. (1993). Cats perceive biological motion. *Psychological Science, 4,* 54–57.

Bonda, E., Petrides, M., Ostry, D., and Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience, 16,* 3737–3744.

Boyce, S.J., Pollatsek, A., and Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 556–566.

Braddick, O. (1974). A short-range process in apparent motion. *Vision Research, 14,* 519–528.

Chatterjee, S.H., Freud, J.J., and Shiffrar, M. (1996). Configural processing in the perception of apparent biological motion. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 916–929.

Cowey, A., and Vaina, L.M. (2000). Blindness to form from motion despite intact form perception and motion detection. *Neuropsychologia, 38*, 566–578.

Cutting, J.E., and Proffitt, D.R. (1981). Gait perception as an example of how we may perceive events. In R.D. Walk and H.L. Pick, Jr. (Eds.), *Intersensory perception and sensory integration* (pp. 249–273). New York: Plenum.

Dannemiller, J.L., and Epstein, W. (1999). Constraining the use of constraints. *Behavioral and Brain Sciences, 22*, 373–374.

Dittrich, W.H. (1993). Action categories and the perception of biological motion. *Perception, 22*, 15–22.

Dittrich, W.H. (1999). Seeing biological motion — Is there a role for cognitive strategies? In A. Braffort, R. Gherbi, S. Gibet, J. Richardson, and D. Teil (Eds.), *Lecture notes in artificial intelligence*, Volume 1739 — *Gesture-based communication in human–computer interaction* (pp. 3–22). Berlin: Springer.

Fodor, J. (1983). *The modularity of mind: An essay on faculty psychology.* Cambridge, Massachusetts: The MIT Press.

Fox, R., and McDaniel, C. (1982). The perception of biological motion by human infants. *Science, 218*, 486–487.

Gregory, R. (1966). *Eye and brain: The psychology of seeing.* London: World University Library.

Grèzes, J., Fonlupt, P., Bertenthal, B., Delon–Martin, C., Segebarth, C., and Decety, J. (2001). Does perception of biological motion rely on specific brain regions? *Neuroimage, 13*, 775–785.

Grossman, E.D., and Blake, R. (1999). Perception of coherent motion, biological motion and form-from-motion under dim-light conditions. *Vision Research, 39*, 3721–3727.

Grossman, E.D., and Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron, 35*, 1167–1175.

Grossman, E.D., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., and Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience, 122*, 711–720.

Haxby, J.V., Hoffman, E.A., and Gobbini, M.A. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*, 223–233.

Heberlein, A.S., Adolphs, R., Tranel, D., Kemmerer, D., Anderson, S., Damasio, A.R. (1998). Impaired attribution of social meanings to abstract dynamic geometric patterns following damage to the amygdala. *Society for Neuroscience Abstracts*, abstract number 463.5.

Henderson, J.M., and Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243–271.

Hollingworth, A., and Henderson, J.M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General, 127*, 398–415.

Hollingworth, A., and Henderson, J.M. (1999). Object identification is isolated from scene semantic constraint: Evidence from object type and token discrimination. *Acta Psychologica, 102*, 319–343.

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

Kosslyn, S.M. (1994). *Image and brain.* Cambridge, Massachusetts: The MIT Press.

Marr, D. (1982). *Vision.* New York: W.H. Freeman.

Mather, G., Radford, K., and West, S. (1992). Low-level visual processing of biological motion. *Proceedings of the Royal Society of London B, 249*, 149–155.

Milner, A.D., and Goodale, M.A. (1995). *The visual brain in action.* Oxford: Oxford University Press.

Neri, P., Morrone, M.C., and Burr, D.C. (1998). Seeing biological motion. *Nature, 395*, 894–896.

Oram, M.W., and Perrett, D.I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. *Journal of Cognitive Neuroscience, 6*, 99–116.

Palmer, S.E. (1975). Visual perception and world knowledge: Notes on a model of sensory–cognitive integration. In D.A. Norman and D.E. Rumelhart (Eds.), *Explorations in cognition* (pp. 279–307). San Francisco: W.H. Freeman.

Pavlova, M., Krägeloh–Mann, I., Sokolov, A., and Birbaumer, N. (2001). Recognition of point-light biological motion displays by young children. *Perception, 30,* 925–933.

Pylyshyn, Z. (1999a). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22,* 341–365.

Pylyshyn, Z. (1999b). Vision and cognition: How do they connect? *Behavioral and Brain Sciences, 22,* 401–423.

Rock, I. (1983). *The logic of perception.* Cambridge, Massachusetts: The MIT Press.

Schyns, P.G. (1999). The case for cognitive penetrability. *Behavioral and Brain Sciences, 22,* 394–395.

Schyns, P.G., and Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5,* 195–200.

Shiffrar, M., Lichtey, L., and Chatterjee, S.H. (1997). The perception of biological motion across apertures. *Perception and Psychophysics, 59,* 51–59.

Sugase, Y., Yamane, S., Ueno, S., and Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature, 400,* 869–873.

Sumi, S. (1984). Upside–down presentation of the Johansson moving light-spot patterns. *Perception, 13,* 283–286.

Thornton, I.M., Pinto, I., and Shiffrar, M. (1998). The visual perception of human locomotion. *Cognitive Neuropsychology, 15,* 535–552.

Thornton, I.M., Rensink, R.A., and Shiffrar, M. (2002). Active versus passive perception of biological motion. *Perception, 31,* 837–853.

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, Massachusetts: The MIT Press.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition.* Cambridge, Massachusetts: The MIT Press.

Ungerleider, L.G., and Mishkin, M. (1982). Two cortical visual systems. In D.J. Ingle, M.A. Goodale, and R.J.W. Mansfield (Eds.), *Analysis of visual behaviour* (pp. 549–586). Cambridge, Massachusetts: The MIT Press.

Vaina, L.M., LeMay, M., Bienfang, D.C., Choi, A.Y., and Nakayama, K. (1990). Intact biological motion and structure from motion perception in a patient with impaired motion mechanisms. *Visual Neuroscience, 6,* 353–369.

Vaina, L.M., Solomon, J., Chowdhury, S., Sinha, P., and Belliveau, J.W. (2001).Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences, 98,* 11656–11661.

Wallisch, P. (2001). The perception of biological motion. Retrieved February 12, 2004, from http://www.lascap.de/Downloads/Biologicalmotion.pdf

Webb, J.A., and Aggarwal, J.K. (1982). Structure from motion of rigid and jointed objects. *Artificial Intelligence, 19,* 107–132.