Mind in Action: Action Representation and the Perception of Biological Motion

Hemeren, Paul

2008

Citation for published version (APA):
Hemeren, P. (2008). *Mind in Action: Action Representation and the Perception of Biological Motion*. [Doctoral Thesis (monograph), Cognitive Science].

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Mind in Action

Action Representationan the
Perception of Biological Motion

Mind in Action

Action Representation and the Perception of Biological Motion

Paul E. Hemeren

LUND UNIVERSITY

Lund University Cognitive Studies 140

Paul Hemeren (2008)
Mind in Action: Action Representation and the Perception of Biological Motion
Lund University Cognitive Studies 140

Cover layout and design: Lennart Danling, Reko Reklam AB

ISBN 978-91-977380-2-6
ISSN 1101-8453
LUHFDA/HFKO 1021

Printed in Sweden by Rolf Tryckeri AB

To Gudrun

Contents

Acknowledgements

There are many people who have contributed in numerous ways to this book. As I reflect on the process of doing research and writing, it becomes obvious that colleagues, friends and family have all played a role. Firstly, I would like to thank my advisor, Professor Peter Gärdenfors, for his keen insights, support and inspiration. After every advising session, I felt a renewed sense of enthusiasm for the research I was doing. The terrific atmosphere of intellectual curiosity and freedom that Peter has fostered has been a hallmark of the Lund University Cognitive Science group, which leads me to mention a few other people who have been a part of that group. A special thanks to Christian Balkenius for creating the software that ran my patch-light displays and for giving me helpful comments on my experiments. Tom Andersson, David DeLeon, Agneta Gulz, Måns Holgersson, Kenneth Holmqvist, Jana Holsánová, Lars Kopp, Robert Pallbo, Annika Wallin and Simon Winter contributed to the lively exchange of ideas during the cognitive science seminars.

I would also like to thank Ian Thornton for agreeing to be the opponent for the dissertation defense and for reading and commenting on earlier versions of the work presented in Chapter 7.

I am grateful for the financial support from Lund University, the Swedish Research council and the University of Skövde. I also received generous financial support in the form of a fellowship from the Knut and Alice Wallenberg Foundation, which allowed me to spend a year at the University of Virginia. Many thanks to Bennett Bertenthal and members of the infant studies lab for hosting my stay.

I have also benefitted from exchanges with collegues and students. In particular, my colleagues Tom Ziemke and Anthony Morse provided valuable feedback on various chapters. I am also grateful to India Morrison for taking time to read and comment on some chapters. Kip Smith also provided much appreciated detailed comments. I've also had valuable discussions with Nils Dahlbäck. I'm thankful for their constructive input and encouragement. The dissertation is better because of it. Thanks to Jessica Linblom and Anna-Sofia Alklind Taylor for discussions about the cover design. I would also like to thank my former students Sigríður Pálsdóttir, Maria Nilsson, Peter Thorvald and Patrik Söderberg for their contributions.

I am grateful to Åse Björnberg and Maria Landström for their assistance with the printing of the dissertation and to Lennart Danling for his creativity with the cover design.

Finally, I would like to express my deep appreiciation to my immediate family for the wonderful support throughout the research and write-up. To my mother Barbara Hemeren for always being a great source of support and encouragement. To my sons Nick and Carl for all the joy I have experienced as a proud father. Lastly, to my wife and soulmate Gudrun, for her unfailing support and love. You mean the world to me. Thank you for being by my side!

List of Publications and Conference Presentations

Hemeren, P. E., Kasviki, S. & Gawronska, B. (2008). Lexicalization of natural actions and cross-linguistic stability. In A. Botinis (Ed.) *Proceedings of the 2nd ISCA Workshop on Experimental Linguistics* (pp. 105-108). Athens, Greece: University of Athens.

Hemeren, P. E. & Gawronska, B. (2007) Lexicalization of natural actions and cross-linguistic stability. In E. Ahlsén et al. (Eds.) *Communication - Action - Meaning: A Festschrift to Jens Allwood* (pp. 57-73). Göteborg University.

Hemeren, P.E. & Gawronska, B. (2007). Cross-linguistic comparisons of spatial patterns for natural action verbs. 10th International Cognitive Linguistics Conference, Krakow, Poland.

Hemeren, P. E. (2005). Orientation specific effects of automatic access to categorical information in biological motion perception. In B. G. Bara, L. W. Barsalou & M. Bucciarelli (Eds.) *Proceedings of the Twenty-seventh Annual Conference of the Cognitive Science Society* (pp. 935-940). Hillsdale, NJ: Erlbaum.

Hemeren, P.E. (2005). The role of high-level representations and conscious awareness in biological motion perception. *Proceedings of ESPP'05, Conference for the European Society for Philosophy and Psychology* (p. 123), Lund University Cognitive Studies.

Hemeren, P. E. (2005). The embodiment of perception: A case for semantic level representations. European conference on Computing and Philosophy, Mälardalen University, Sweden.

Hemeren, P. E. & Pálsdóttir, S. (2003). Category effects and orientation specificity in inverted and upright displays of biological motion. *Perception*, 42, 142.

Hemeren, P. E. (2003) Long-term priming and orientation specificity in the perception of biological motion. In T. Bajo & Lupiáñez (Eds.) *XIII Conference of the European Society of Cognitive Psychology* (p. 250). Universidad De Granada.

Hemeren, P. E. (1997). Typicality and context effects in action categories. In M. G. Shafto & P. Langley (Eds.) *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (p. 949). Hillsdale, NJ: Erlbaum.

Hemeren, P. E. (1996). Frequency, ordinal position and semantic distance as measures of cross-cultural stability and hierarchies for action verbs. *Acta Psychologica*, 91, 39-66.

Hemeren, P.E. (1993). Actions and biological motion: A method for generating dynamic displays in experimental settings. Paper presented at the Primacy of Action : An advanced inter-disciplinary workshop, Manchester, England.

Chapter 1 - Introduction

The beauty of human movement shows itself in a variety of ways, from the well defined graceful movements in a ballet performance to a perfectly executed double play in baseball. In both cases, beauty may reside in the eye of the beholder, or author. Knowledgeable observers and professional performers have developed an ability to recognize fine-grained patterns of human movement. Even less knowledgeable observers can appreciate the skillful motoric artistry in various sporting events and performing arts. Appreciation, or recognition of human motion, however, is not restricted to professional performers or expert observers.

On a more daily basis, an important aspect of human cognition is the ability to perceive and understand the actions of other individuals. In this sense, appreciation of human movement is defined less by artistic convention than by a basic need to interact with things around us and other individuals. The importance of this ability is reflected in basic survival value as well as socially oriented situations. Perceiving the difference between threatening and friendly behavior allows us to avoid harm and to seek out socially beneficial contact. The ability to recognize the actions of others allows us to adjust our actions accordingly. For example, when approaching a friend I have not seen for a while and he stretches out his open hand towards me, it clearly means that I should shake his hand in a friendly way. Within the area of coordinated activities like playing sports, the ability to perceive the actions of others as a kind of prelude to what is going to happen next is crucial. The role of dynamic information in the perception of actions could be an important factor that distinguishes action categories from, for example, artifact categories like *houses*, *cars*, *books*, *pencils*, etc.

When we see the movement of other individuals, we do not merely see the independent movement of arms and legs connected in a specific way to the torso. We instead see meaningful actions like waving, running, crawling, swimming, etc. Furthermore, this perceptual/cognitive ability is done without ‘thinking’ about how we do it. It is seemingly effortless in many cases. Given the role that this ability plays in our everyday cognition, it seems important to investigate how this ability arises. This book is about our ability to recognize and categorize human actions.

It is no small secret that human vision is highly sensitive to the motion patterns created by the movement of other individuals. When we identify a motion pattern as an instance of someone running or walking, we are categorizing that motion pattern. This ability to see a pattern of human motion as an instance of running or walking and not just as a complex pattern of movement of the arms and legs is sometimes referred to as epistemic visual perception (Jeannerod & Jacob, 2005). A further aspect of the sensitivity of human vision is that we are able to see the intentions of others in their basic actions (e.g., Blakemore & Decety, 2001; Dittrich & Lea, 1994; Iacoboni et al., 2005; Saxe, Xiao, Kovacs, Perrett & Kanwisher, 2004). The meaning of actions is therefore determined by the conceptual knowledge associated with a given pattern of bodily motion. Conceptual knowledge in turn includes knowledge about the goals, emotions, sensory-motor patterns, intentions and associations to related action categories in an action hierarchy.

The empirical findings in this book contribute to our understanding of how we can perceive the actions of other humans. More specifically, the findings expand upon previous research about the role that high-level conceptual knowledge seems to play in the structure of action categories and our ability to recognize the actions of others. The high-level conceptual knowledge referred to here will focus on the role that sensory-motor patterns and associations to related action categories play in action recognition and categorization. An additional aspect of the research presented concerns the relationship between perception/conceptualization and the words used to express what we see. To what extent do the associations between the words we use reflect the way they are cognitively organized?

In addition to introducing the central questions covered in this book, this introductory chapter will present the following:

- the embodied approach to cognition, which serves as a theoretical context for my research
- the development of the research topics
- a cognitive science perspective
- a description of the chapters
- a summary of the main thesis and contributions
- paths not taken.

1.1 An Embodied Perspective

Although I will not directly investigate the embodied approach to cognition (e.g., Barsalou, 1999; Chrisley & Ziemke, 2003; Clark, 1999, 2006; Svensson, Lindblom & Ziemke, 2007), I will simply point to the fact that numerous results support the idea that action representation, recognition and categorization critically involve the ability to mentally simulate and relate the actions of others to one's own body and action repertoire.[1] Mental (motoric) simulation of the actions of other people seems to be intricately linked to understanding observed actions (e.g., Blakemore & Frith, 2005; Calvo-Merino, Grèzes, Glaser, Passingham & Haggard, 2006; Casile & Giese, 2006; Jeannerod, 1994; Sebanz, Knoblich, & Prinz, 2003; Shiffrar, 2006, 2008). The mental simulation referred to here should not be confused with visual mental representation in the form of mental images. Instead, mental simulation should be understood as an actual motor based spatiotemporal representation of an action. Findings from Lozano, Hard and Tversky (2007, 2008) have also shown that peoples' understanding and descriptions of objects in a scene are influenced by their own body and motor representations. The extent to which people are able to relate to their own motoric capabilities affects their comprehension of the interaction with objects.

The role of mental simulation in action understanding is also apparent in a cognitive impairment known as apraxia. According to Jeannerod (2006), a central role of motor simulation is apparently lacking in apraxia, in which a person has difficulty in performing skilled actions usually requiring the use of a tool. In addition to this difficulty, apraxia results in an impairment of the ability to pantomime common uses of tools, like hammering, cutting paper with a pair of scissors, etc. Together with these deficits, action simulation and action recognition are also impaired. It is not the case, however, that people with apraxia lack the ability to reach and grasp objects, which means that the apraxic impairments are not due to a pure motor or visual deficit (Jeannerod, 2006). The impairment seems to be due to an inability to select the appropriate motor elements that figure into a goal directed action (Jeannerod, 2006).

Recently, much research has demonstrated a central role for the involvement of motor resonance in language understanding, especially the understanding of verbs for concrete bodily actions (e.g., Arbib, 2008; Aziz-Zadeh & Damasio, 2008; Gallese & Lakoff, 2005; Hauk, Johnsrude & Pulvermüller, 2004; Tomasino, Fink, Sparing, Dafotakis & Weiss, 2008; Tomasino, Werner, Weiss & Fink, 2007). Hauk et al. (2004) used functional magnetic resonance imaging (fMRI) to show that verbs referring to face-, arm-, and leg-related action during a passive reading task lead to significant levels of cortical activation along the motor strip. The areas activated along the motor cortex were very close to or directly overlapped with the areas that are activated when we use the tongue, fingers or feet. In short, there was a somatotopic activation of the premotor and motor cortex. Recently, Tomasino et al. (2008) found that when transcranial magnetic stimulation (TMS) was applied to the hand area of the left primary motor cortex, facilitation occurred for the processing of verbs related to hand actions. This facilitation occurred when subjects used motor imagery to process

---

[1] See e.g. the anthologies edited by Klatzky, MacWhinney and Behrmann (2008) and Ziemke, Frank and Zlatev (2007) for overviews of the issues in the embodied approach to cognition. The anthology "Common mechanisms in perception and action: Attention and performance" (Prinz & Hommel, Eds., 2002) also contains numerous articles that present summaries documenting the cognitive importance of the perception-action connection.

hand-related action verbs. It should be pointed out, however, that while motor imagery has a facilitating effect and often occurs as a result of processing the meaning of action verbs, Tomasino et al. (2008) found no evidence for the *necessity* of motor imagery for action verb processing.

Despite the wealth of new behavioral and neuroscientific evidence in support of the close relationship between action observation and understanding, there is an ongoing debate about the extent to which action understanding in particular and human cognition in general is "embodied." Mahon and Caramazza (2008) attempt to frame the behavioral and neuroscientific results that support an embodied view in terms of a middle ground between a "pure" embodied approach at the one extreme and a "pure" disembodied hypothesis approach. According to their view, sensory and motor information are necessary for *online* conceptual processing but not necessary for a conceptual level that is more abstract and symbolic. For further information about the specific issues in this debate, see Mahon and Caramazza (2008) and Shapiro, Moo & Caramazza (2006).

In this short summary of the embodied approach to cognition, I should mention the role that the discovery of mirror neurons has had for the development of embodied cognition. Briefly, mirror neurons become activated when an individual performs certain actions *and* when an individual observes the actions of another person performing the actions (e.g., Buccino, Binkofski & Riggio, 2004; Rizzolatti, Fadiga, Gallese & Fogassi, 1996; Rizzolatti, Fogassi & Gallese, 2006). This discovery first occurred in monkeys (di Pellegrino, Fadiga, Fogassi, Gallese & Rizzolatti, 1992), and current research has focused on finding similar populations of neurons in human subjects using brain imaging techniques.[2] The connection to the embodied cognition approach is obvious. Mirror neurons seem to fill the cognitive processing gap between visual input and motor output in, for example, imitation (Meltzoff & Moore, 1977; Rizzolatti, Fogassi & Gallese, 2001; Wilson, 2006). Prior to the discovery of mirror neurons, the imitative capacities in, for example, neonates was thought to be explained by an intermodal matching process (Meltzoff & Moore, 1983). With the discovery of mirror neurons, there is no need to posit a matching process. The neurons needed to produce the actions are directly activated through observation. To the extent that knowledge is gained through and structured by the interaction of the human body with the physical environment, including other human bodies, the contribution from the discovery of mirror neurons is central to the embodied cognition approach.[3]

---

[2] See e.g., Rizzolatti (2004) and Rizzolatti & Craighero (2004) for reviews.

[3] I should also note that the claimed explanatory breadth of mirror neurons in human cognition has its detractors. For a critical analysis of the role of mirror neurons in humans, see e.g., Turella, Pierno, Rubaldi and Castiello (2008).

1.2 The Development of Research Topics

The course of development for the ideas in this book have been shaped by what I perceived to be areas in which research was lacking and could potentially constitute a scientifically fruitful path of investigation. My original research interest was in the area of categorization. I soon found that much of the research at the time (mid 80s) was focused largely on static objects (natural kinds and artifacts). My impression is that this is still pretty much the case. I thought this emphasis on static objects neglected an important aspect of our daily activity, namely, our ‘dynamic’ interaction with our physical and social environment, including our ability to recognize the actions of other individuals. Therefore I began to look at how I could study action concepts with the purpose of seeing if action concepts were psychologically organized in ways similar to object concepts, or categories. I soon discovered one reason why researchers may have neglected studying action concepts. In addition to the fact that action concepts are difficult to define, they were, at that time, difficult to use as well controlled stimuli on a computer. This technological limitation was overcome by adopting the point-light technique as first used by Gunnar Johansson (1973) in experiments on biological motion perception. (The point-light technique and its current applications are presented and discussed in Chapter 4.) The gist of the technique is to capture the motion of the human body by filming the motion of the joints. This was done by putting small light bulbs or reflecting markers on the joints of a human actor. Since the motion of the points of light conveyed information about the movement of the human body and all other information was filtered out, the resulting animations could be easily displayed out on a desktop computer. The technique allowed me to present actions in real-time on a computer as well as providing a second benefit, namely, a way of controlling extraneous variables like body shape and other contextual factors that would easily confound the experiments I was contemplating.

I also became aware that I was looking at two research areas at the same time. One area was categorization and the other was biological motion perception. These two areas are represented in this book. The broad purpose of this book is to relate the areas of action categorization and action perception to one another. This broad purpose, in turn, consists of two prongs of investigation. The first is to investigate action categories from the perspective of previous findings within categorization research. To what extent do action categories exhibit “classical” findings concerning hierarchical structure, basic level effects and graded structure in reference to an action prototype? The second purpose is to investigate the perception of actions using point-light displays of biological motion. This latter purpose attempts to relate the activation of categorical information to the visual processing of actions in displays of biological motion. In the same sense that categorization research has been silent about the domain of action categories, the literature on biological motion processing has, until fairly recently, been similarly silent about the role of categorical knowledge in the visual processing of biological motion. Of course there are exceptions to both domains of silence, and those exceptions will be discussed further in Chapters 2 and 6. The goal of the research presented here is to gain a better understanding of the categorical/conceptual knowledge associated with actions and the possible role of this knowledge in the visual processing of actions. How do we recognize the actions of others? I propose that one important aspect of the process of recognition has to do with the activation of categorical knowledge in the form of stored spatiotemporal patterns of human movement that are organized according to action prototypes.

Previous research and models within action perception using displays of biological motion have suggested a clear role for categorical knowledge of actions. Little work, however, has been done to specifically investigate the categorical structure of action categories. Dittrich (1999), however, proposed a sketch of a model (Interactive Encoding Model) where high-level categorical knowledge is proposed as playing a central role in the perception of biological motion. He argues for the existence of a functional route in biological motion processing that strictly relies "on visual-semantic information which is stored in respect to action categories" (p. 16). In other words, Dittrich (1999) argues for conceptually driven processing in biological motion perception.

The core of Dittrich's (1999) interactive model consists of three functionally specified routes that deal with the integration of motion information. One route appears to be specified by the use of 2D information to recover the 3D form of the human body. A second suggested route processes information about the constraints, or built-in assumptions, of the possible paths of human movement. The effect of the built-in assumptions therefore relies on a link to a memory system that contains information about motion paths related to motion categories. Semantic level effects could include representational momentum, which refers to our ability to represent the paths of objects beyond what is directly given in the visual stimulus (Shiffrar & Freyd, 1993). The third route, according to Dittrich (1999) deals with the processing that involves visual-semantic information that is characterized by the stored knowledge of action categories. One prediction based on the processing in this route is the occurrence of prototype effects. Access to the meaning of human body motion can be affected by perceptual matching to stored exemplars on the basis of the goals and intentionality associated with human movement. A key feature of Dittrich's idea of interactive encoding is that even access to semantic level information may occur early on in the visual processing of biological motion. Indeed, as Dittrich asserts, "Visual processing always appears to involve some kind of intentional aspect. It is a design feature of the visual system" (p. 18).

In their hierarchical model of human recognition of biological movement, Giese and Poggio (2002, 2003) also propose that both the form (ventral) and motion (dorsal) pathways contain high level areas that represent action specific information as action prototypes. It is difficult on the basis of their proposed model, however, to determine the extent to which categorical knowledge of actions is also stored together with these motion pattern neurons. There is reason to believe that this is not the case. Since their model is based on strictly feedforward processing, effects of top-down constraints due to the activation of categorical knowledge will not be seen in the model. However, some constraints regarding categorical knowledge will result from the activation of action prototypes. One central question here is what other high level association areas are involved in action recognition that are not a direct result of access to prototype representations of specific actions. Such association areas may be involved in determining the goals and intentionality of actions, as well as perceiving the intentions of others in the actions they perform.

An important finding in categorization research is the existence of a basic level at which (static) objects seem to be categorized. Similar to the findings supporting basic level categorization for objects, it will be necessary to find converging evidence for the basic level for action categories. Results from previous research show that there is converging evidence for a basic level for objects. This converging evidence comes from different areas such as feature listing/similarity judgments, motor routines used in the interaction with objects, the visual form of objects, category membership judgments, word use and word structure (Murphy, 2002). This book presents some findings that have some bearing on the issue of a basic level for action categories. A further contribution regarding a basic level for action categories is to relate findings from the basic level for object categories with the purpose of generating research issues about the psychological organization of action categories. The purpose is to pose the important question about the structure of action categories and the role that categorical knowledge may play in the perception of actions. These two issues can be understood as two different approaches to investigating the organization of categorical knowledge of actions. One approach is to investigate the structure of action categories from the field of categorization. The methods used here include previous methods used in categorization studies of objects where different levels of categorization are used to see if processing differences occur as a result of the different levels. This has been the tradition within much of the categorization research. Another approach is to investigate the perception of actions using psychophysics and try to see to what extent categorical knowledge may be used to recognize or identify actions. The inherent strength in including the two approaches can be seen in the attempt to integrate research from categorization and perception, as well as to take relevant neuroscientific results into consideration. An additional purpose is to find and generate new research issues and applications.

1.3 A Cognitive Science Perspective

Being a book in the area of cognitive science, the work presented here is interdisciplinary. It covers a methodological spectrum from psycholinguistics to psychophysics as well as considering recent results from cognitive neuroscience. The potential problem that arises in this broad interdisciplinary context is one of maintaining a scientific balance between broadness of scope and detailed, well-controlled experimental studies. In metaphorical terms, the challenge is maintaining a broad focus on the ‘big picture’ while investigating the smaller pieces of the larger puzzle of human cognition.

An important aspect of this book is to relate different areas of cognitive science to one another through the question of how we can talk about what we see (Jackendoff, 1987). In the case of this book, the object of what we see is restricted to the actions of others. The ability to understand and communicate about the actions of others is a fundamental aspect of our daily activity. How can we talk about what others are doing? What qualities do different actions have such that they cause us to see them as being different or similar? What is the connection between what we see and the development of concepts and words or expressions for the things that we see? To what extent can two different people see and talk about the same things? Is there a common basis for our perception, and is there then a common basis for the concepts we form and the way in which the concepts become lexicalized in language? What influence do social and cultural aspects have on our perception and categorization of actions, and is this potential influence “visible” in different languages?

Although many of these questions form the context for future research and will not be a part of the puzzle pieces in this book, the description of the theoretical background as well as the empirical studies attempt to relate language, categorization and perception. The details about the specific relations between language, categorization and perception are addressed to some degree but also serve to generate further research issues.

1.4 Description of the Chapters

There are three themes to the book. These themes correspond to the topics mentioned previously, i.e., language, categorization and perception. Each theme consists of a chapter that presents the previous studies within the area followed by a chapter that presents my own empirical studies. This is the pattern of presentation for the three themes. The empirical studies in Chapters 3, 5 and 7 have all either been published in a journal or in conference proceedings. The references to the journals and conference proceedings are given at the beginning of each chapter. Chapters 2, 4, 6 and 8 have not been previously submitted for peer-review, although the material in Chapter 4 has been included in a conference presentation.

Chapter 2 reviews the categorization literature with an emphasis on evidence for the existence of concept hierarchies, prototype effects and the basic level for static objects. The purpose of this chapter is to relate previous findings from categorization research to the domain of action recognition and identification. How are action categories cognitively organized in relation to possible different levels of abstraction? Do action categories exhibit prototype and basic level effects? Given the relatively large amount of categorization research, it seems a reasonable starting point to use those results to pose questions and construct experiment using actions as stimuli rather than static objects. Many of the issues raised in this chapter provide the basis for the empirical investigations in later chapters, especially Chapters 3 and 5. The expression of actions in language through the use of verbs for natural actions is also addressed. Differences and similarities between the cognitive organization of nouns and verbs are described as a way of understanding possible representational differences between static objects and natural actions. Finally, recent models for action representation and recognition are discussed.

Chapter 3 contains two empirical studies that investigate the relation between perception and the hierarchical structure of action categories, i.e., subordinate, basic, and superordinate level action categories. Issues regarding the cross-cultural stability of the cognitive organization of action categories are also addressed by including American English and Swedish speaking subjects in a verb listing task. The first study includes the American English speaking subjects, and the second study includes the Swedish speaking subjects. Analyses of the list data are presented separately for each group as well as a cross-linguistic analysis. Multidimensional scaling is used to assess the potential overlap between the two language groups. Results that show a strong tendency for both groups to list very similar actions, e.g., run, jump, walk, kick, swim, scream, eat, cry, etc., are presented. Further results are discussed in relation to basic level and prototype effects for action categories.

Chapter 4 describes the development of the point-light technique that is used to create displays of biological motion. The point-light technique has been used in many experiments since Gunnar Johansson (1973) first used it in his experiments on human action perception. This chapter describes the general technique as well as specific developments and applications within current research. I also describe how the technique has been specifically adapted to the experimental settings discussed in Chapters 5 and 7.

The empirical studies in **Chapter 5** investigate the extent to which action categories exhibit graded structure, which would indicate the existence of prototypes for action categories. This issue is addressed in two experiments. Are the results from action categorization studies consistent with previous categorization findings from the domain of static objects? How might action categories differ from other categorical domains? In the first experiment subjects are asked to rate the typicality of different kinds of actions presented as point-light displays. The results from the typicality ratings are then assessed in order to determine potential typicality differences among the actions. These results are then compared to the results from the second experiment which, in contrast to the first experiment, uses a category verification task. For both experiments, the subjects see the same point-light actions, and the major question concerns the extent to which typicality judgments can be used to predict category verification times for instances of different actions.

Chapter 6 provides a review of the research on the role of attention and levels of processing in biological motion perception. The issue of the extent to which biological motion perception is dependent on the orientation (orientation specificity) of the point-light displays is also discussed. Previous research has shown that action recognition is impaired when point-light actions are view up-side-down, which suggests that the top-down processing of configural information plays an important role in the visual processing of point-light displays of biological motion. On the basis of the results from the previous chapters about the structure of action categories, I raise the possibility that information about action categories is available to subjects and that category knowledge can be activated implicitly.

Chapter 7 presents an experiment based on a repetition (short-term) priming paradigm. The experiment investigates the question of whether or not the visual processing of biological motion displays includes high level information about the categorical differences between actions. The extent to which access to categorical differences differs as a function of display orientation is also investigated. The results are analyzed and interpreted within the context of the previous research presented in Chapter 6 and within the context of the more specific questions posed at the beginning of the chapter.

Chapter 8 includes a recapitulation of the central issues in the book. A summary of the main findings and contributions are also presented. The main findings are discussed within the broader theoretical context mentioned Chapters 1 and 2. Finally, I discuss concrete proposals for further empirical studies and relate some of the findings to issues in artificial intelligence and information technology.

1.5 Main Theses and Contributions

The previous sections have addressed many of the issues that are covered in the research presented here. For the sake of clarity, I will present a brief summary of the main theses and contributions.

The major theses are:

- We organize our knowledge about our actions and the actions of others in the form of hierarchically structured dynamic action templates or prototypes. We have access to, and use, this information even when we perform a perceptual task that does not require it. The existence of hierarchically structured dynamic action prototypes can also be seen in results from cross-linguistic comparisons.
- Cognitive access to dynamic action prototypes requires visual configural processing and is orientation specific, i.e., limited to canonical[4] upright displays.

---

[4] The term ‘canonical’ is used to allow for the possibility that some upright displays might not be usual or familiar, e.g., walking on hands (Shipley, 2003). In this case, the canonical orientation is the one most often seen, i.e., with the observer upright and the actor upside down.

The contributions consist of the following:

- One contribution is methodological. This contribution consists of using multidimensional scaling (MDS) to convert verb list data into semantic distances which can then be used as a basis for comparing the semantic spaces for verbs across languages. I should point out that the claimed contribution is *not* the development of the MDS technique but rather the application of MDS in order to create a more fine-grained representation of the relationship between action verbs in American English and Swedish. To my knowledge, this is the first application of MDS to cross-linguistic studies of action verbs.
- The results from the experiments show a clear effect of the graded structure of the action categories included in the experiments. Action categories, under some circumstances, have prototypes. In addition, I present evidence for action hierarchies. The results from these experiments indicate that people have access to information about action meaning in the sense of making categorical distinctions between actions.
- A further contribution to the area of biological motion perception is the finding of an implicit activation of categorical level processing for actions and that this information is available for upright displays, but appears to be lacking for displays that are shown upside down. The previous research on the orientation specificity of biological motion perception has shown that the visual processing of upside down displays is *not* facilitated by first viewing an upright display. However, the reported results in Chapter 7 show this facilitation can be obtained. In this case, the contribution consists of obtaining results that to some extent go against previously obtained results.
- Given these contributions, the next question concerns their implications. This will be discussed in the last chapter of the book.

1.6 Paths Not Taken

In this section, I would like to mention some of the potentially relevant areas that I have purposely chosen to avoid in the book. For example, I will not be discussing developmental aspects of biological motion processing. The relevant theoretical background and results from previous research can be adequately described without going into the developmental literature. The interested reader is referred to the informative overview by Pinto (2006).

What is an action? Since a philosophical discussion about what properly constitutes an action is of little relevance to the work presented here, I will refrain from an attempt to precisely define what constitutes an action. The notion of an action as it pertains to the empirical work presented here is perhaps best described by examples. Running, jumping, swimming, waving, kicking, talking, throwing, etc. are motor-based actions. Jackendoff (1990) refers to such actions as “natural actions.” Based on the work of Peterson (1985), Jackendoff suggests that natural actions are difficult to describe but easy to point out. Humans appear to be able to identify actions at this level of description. We also appear to communicate our own actions and the actions of others by using words to describe human movements on that level. This should not be taken to mean that no other level of description or communication is used in understanding the actions of others. We know for example that when actions are associated with a specific goal, this can influence the way we perceive and communicate our understanding of our own actions as well as the actions of others (Vallacher & Wegner, 1987). For example, the action of running may not simply be seen as running but rather as trying to scare away a stray cat that has come into the yard. It is certainly the case that the goal structure of actions influences our categorization of human movement. In this sense, actions can be highly context sensitive. The investigation of actions in this book is based, however, more on the systematic differences in perceptually given spatiotemporal aspects of human movement. While acknowledging the role of goals in action perception, I will not be investigating the systematic effect of goals on our perception of actions. The emphasis is rather on the role of the spatiotemporal patterns of human movement in action perception.

One aspect (philosophical) that does play a role in what I refer to as action is intention. The actions to which I refer all have an intentional component. I assume that a person performing an action can generally be ascribed the intention of carrying out that specific action. Another aspect has to do with the scope of the actions addressed here. The research here is largely confined to motoric actions, as suggested in the examples above. These actions can generally be recognized within short time frames, roughly around 300 milliseconds (Johansson, 1973).

There is also a relevant distinction to be made between actions and events. The distinction is pragmatic and not intended to be a logically well-defined description of the difference between events and actions. Although there is some overlap between them, I want to avoid potential confusion between the two terms. For the sake of simplicity, the most salient difference is that events do not necessarily entail human action. A beautiful sunrise is such an event. And in this case the distinguishing factor is the absence of an intentional agent. Despite the clarity of this example, there are cases where is it difficult to draw a clear line between an event and a complex action. Is ‘answering the telephone’ an event or an action? In this case, I am prepared to say that it is both, in the sense that there is a clear intentional goal (to respond to someone’s request to speak with you), and there is a fairly clear motor component, namely lifting and speaking into the receiver. To the extent that this book deals with action perception and categorization, I will refer to such examples as actions, not as events. That is not to say that event perception and the perception of actions have no common basis. Nor do I claim that research in the one area is not relevant to the other.[5]

In contrast to actions, events represent a potentially larger unit of analysis. For example, buying groceries is an event that consists of various actions like reaching, grasping, walking and talking. The event also importantly includes interaction with objects as in picking up, putting down, pushing a cart and paying. It could be argued that buying groceries can also be viewed as a complex action consisting of a number of constituent actions. Another important factor that can differentiate the actions discussed here and the notion of events is the apparent goal structure of events. Eating at a restaurant (Schank & Abelson, 1977) can be described as consisting of 4 subgoals (or scenes): entering, ordering, eating and exiting. The action of 'running' however may be influenced by different goals, but its identification appears to be less influenced by the goal structure in which it occurs. Put another way, while the spatiotemporal pattern of running can figure in a number of different events like running a race, running to catch a train, chasing away a cat, etc., it is much more difficult to think of a context where the spatiotemporal pattern of running would *not* be seen as running. In this case, I suggest that the motion as such is not (or a least much less) goal defined, you do not need to know the context sensitive goal in order to identify the action. At least it is not goal defined in the same sense as human activity oriented events, which seem to require a goal structure. This distinction is likely more a matter of degree than kind. Consequently, I will be addressing issues of categorization that deal with "natural actions" rather than human activity based events.[6] For a more detailed description of event structure see, for example, Zacks and Tversky (2001), Newtson (1973), Newtson and Engquist (1976) and Newtson, Engquist and Bois (1977).

Some actions take longer to perform than other actions. Some are cyclical in nature (walking, swimming and running) and others are non-cyclical, e.g., throwing, kicking a ball and sneezing. Actions can also differ according to complexity. The notion of complexity can be approached from a number of different perspectives. Motoric complexity (walking vs. pirouette), the participation of objects in actions (throwing a baseball vs. opening an umbrella), the participation of other individuals in an action (dancing vs. wrestling) temporal extent (catching a ball vs. dribbling a ball) represent different approaches to "parsing" or understanding the structure of actions.

---

[5] See Zacks and Tversky (2001) and Zacks (2004) for related literature on event structure in perception.

[6] I should point out here that I am well aware of the role that goals play in the perception of events and action planning. Hommel, Müsseler, Aschersleben and Prinz (2001) present an in-depth analysis of goals in their Theory of Event Coding. The work in my book has not addressed this important aspect of action perception and action planning and the action-perception coupling. I have instead chosen a "narrower" focus by investigating action perception and categorization via point-light displays of biological motion.

All of these distinctions are relevant to the issues being addressed in this book. For the sake of being able to reach interpretable conclusions, I have tried to limit the dimensions by which different actions can vary. I discuss this issue to a greater extent in Chapters 2 and 4.

Chapter 2 - Concepts, Categories and Actions

Action categories are just one group among a myriad of categories and concepts that humans possess. Before discussing issues specifically related to action perception and categorization, it is necessary to present a more general background to research on concepts and categorization. This general background, however, will not consist of a summary of the field of categorization and concept acquisition.[7] The purpose here is rather to present the theoretical and empirical background that is relevant to the issues addressed in this book. To that end, the chapter is structured around those issues. But before delving into the background proper, I will briefly clarify what I mean by the following terms: *concept*, *category* and *categorization*.

To begin with, the notion of a *concept* is highly problematic. Issues of what constitutes a concept and what the functions of a concept are have been the object of philosophical and psychological work since antiquity (e.g., Fodor, 1998; Gärdenfors, 2000; Barsalou, Simmons, Barbey & Wilson, 2003). For the work presented here, I will simply, and perhaps controversially (see Medin, Lynch and Solomon, 2000), use the term *concept* to refer to a mental representation that contains knowledge about an object or class of objects that serves to pick out or point to the object or class of objects that are characteristically associated with the concept. Two points should be noted here. The first is that the idea of an object is broadly defined to mean any entity or phenomenon (or classes) that can be characterized according to stored or directly perceived knowledge about the entity or phenomenon (or relevant classes). The second point concerns mental representation. There is no implied suggestion in the proposed definition as to how concepts are represented *mentally*, i.e., as discrete or distributed mental representations. Related to this is the issue of how concepts are neurologically instantiated in the brain. I make no explicit or implicit claims about

---

[7] For readers interested in such a summary, see Murphy's (2002) *The Big Book of Concepts*. For a collection of original papers collectively representing interdisciplinary aspects of categorization see Margolis and Laurence (1999) *Concepts: Core Readings*. See even Komatsu (1992) for a brief review.

how concepts are instantiated in the brain. This issue, however, is different from understanding how the brain processes *information* associated with concepts. One important aspect of studying concepts is investigating how conceptual knowledge is obtained and processed according to psychological and neuropsychological principles. To this end, findings from psychology and neuroscience will be presented where relevant.

In contrast to *concepts*, *categories* refer to a partitioning or class of objects (entities) that have been grouped together according to relevant characteristics or properties. *Categorization* refers to the *process* whereby ‘objects’ are grouped together based on some commonly shared properties. Determining what kind of information is used to group ‘objects’ and how the grouping process(es) functions represent central issues in categorization. The work presented in this book does not systematically investigate the issue of different kinds of information used in the categorization process in order to draw general conclusions about that issue. Nor is the intention to investigate the *general* issue of how categorization processes function. The work presented here, however, does specifically address the information used to *categorize actions*, in particular actions presented in point-light displays of biological motion.

2.1 Concepts and Conceptual Knowledge

As a background to action perception and categorization, consider the views expressed in the following two quotes:

> “Concepts are the *glue* that holds our mental world together.” (Murphy, 2002, p. 1, italics added)
>
> “Without concepts, there would be no thoughts. Concepts are the basic *timber* of our mental lives.” (Prinz, 2002, p. 1, italics added)

These quotes reflect the fundamental nature of concepts as elements of mental structures. Although the metaphors of ‘glue’ and ‘timber’ suggest slightly different perspectives on the nature of concepts, they similarly point to the role of concepts as somehow organizing knowledge that we have about things in our surroundings. Concepts allow us to communicate ideas, make inferences and understand what is happening around us. They are important in thought and communication. The ‘things’ that concepts refer to include objects (artifacts and natural kinds), places, people, biological and social relations, food, music, emotions, language, faces, events and actions. Our knowledge about ‘things’ also includes abstract concepts like democracy, beauty and truth. We also have concepts for sensory qualities like smells, tastes, sounds, touch, colors, textures, etc. The variety of concepts that we possess speaks to the many ways that knowledge can be organized and used. It also indicates our ability to form new concepts based on new information. When confronted with a situation not previously encountered we can even use ‘old’ stored information to understand and adapt to the new situation.

The notion that we have concepts and that concepts represent organized knowledge raises the issue of how knowledge is structured or organized by concepts. What are the factors or principles that contribute to the formation of concepts? What knowledge do we associate for example with the concept of DOG?[8] Surely, a dog has four legs, barks, has fur, ears, a tail etc. But there is also other knowledge associated with our concept of DOG. Dogs can be used to hunt, guard, rescue, race, pull things and play with. They can make good companions and like to go for walks. While our DOG concept contains knowledge about the physical appearance of dogs, other knowledge is more functional, i.e., dogs have a function for us as human beings.[9]

If concepts are the ‘glue’ that holds our mental world together, then we might want to know what it is about the ‘glue’ for individual concepts that holds some knowledge together but rejects other knowledge as irrelevant. In other words, why are we inclined to use certain properties to identify, say, a dog but use other properties to identify, say, a chair? This is known as the coherence aspect of categorization (Komatsu, 1992). The coherence aspect also addresses the coherence of categories, i.e., the fact that creatures that have dog-like properties are grouped together and are picked out by the DOG concept.

One reasonable approach to category coherence is to suggest that perceptual or sensory qualities play an important role when applying concepts in order to group or classify things. One reason why dog-like creatures are grouped together is that they all (with some exceptions of course) share certain physical qualities. There is a degree of perceptual similarity between dog-like creatures that allow us to group them together. They have fur, four legs, bark, growl, etc. So, perceptually based information appears to be an important source of information in our representation of concepts, at least for concrete ‘objects’. Consequently, this kind of information also likely plays a role in the categorization of objects. But let us take a slightly different look at the role of perception in conceptual knowledge.

2.1.1 Perceptual Symbol Systems and Conceptual Knowledge

To say that perceptually determined features figure prominently in our conceptual knowledge can be interpreted as meaning that the role of perception is restricted to providing perceptually determined bits of information to a conceptual system that also may consist of other forms of “non-perceptual” information like functional and causal

---

[8] I will use upper case letters to refer to concepts and italics to refer to categories. Italics will also be used periodically for emphasis, the use of which should be understood by the context.

[9] Carl von Linné (1756) provides a description of races of dogs based on their domesticated qualities. This work of Linné’s (Cynographia eller Beskrifning om Hunden) was reprinted as ”En Gammal Svensk Hundbok” (An Old Swedish Dog Book) in 1962 (Bokförlaget Fabel, Sigtuna).

relational information (e.g., Medin & Ortony, 1989; Smith & Medin, 1981; Mandler, 2004). In contrast to this way of viewing the role of perception in categorization or human cognition, Barsalou (1999), Barsalou et al. (2003) and Prinz (2002) propose a more pervasive role for perception.

According to their view, conceptual knowledge is grounded in perception. Perceptual systems are themselves representational and not merely information servants that feed the conceptual system with sensory-based information. Barsalou’s (1999) theory of perceptual symbol systems is a theory of knowledge, not of perception. The basic gist of the theory is that all conceptual knowledge is a result of processing by sensory-motor mechanisms. Barsalou’s theory takes this assertion even further. Not only is conceptual knowledge a result of *processing* by sensory-motor mechanisms, but conceptual knowledge is *represented in* sensory-motor areas of the brain. The implication of this is that human cognition, to the extent that it reflects conceptual knowledge as mentioned above, is largely determined by sensory-motor systems. Perceptual symbols then according to Barsalou are the neural activations/representations that correspond to sensory-motor interactions (real or simulated) with our environment and with other concepts that we possess. In this sense, human cognition is not just limited to immediate interactions with our environment. Because we have stored conceptual knowledge in long-term memory that is also represented in sensory-motor areas, we are able to entertain thoughts in the absence of an immediate sensory stimulus. This ability allows us to make plans and simulate their possible consequences without having to actually perform the steps in carrying out the plan. It also allows us to deal with the present situation and reconstruct past events (cf. Hesslow, 2002; Grush, 2004).

Grounding conceptual knowledge in sensory-motor representations leads to the more specific view that conceptual knowledge is also modality specific. In this case, modality specific conceptual knowledge about the visual appearance of an object is represented in cortical areas that process visual information. Similarly, conceptual knowledge about the sound of a barking dog is represented in cortical areas that process auditory information. It also means that the function of an object is represented in cortical areas that process somatosensory information and likely visual information if the function of the object is partly mediated by vision.[10]

One important further implication of Barsalou’s theory of perceptual symbol systems is that the distinction between perception and cognition becomes blurred. If cognition itself is based on modality specific perceptual symbols rather than amodal or abstract symbols, then cognition is grounded in perception. Given his theoretical framework, perceptual and conceptual processes are intricately linked; “Because

---

[10] See for example Barsalou (1999) for evidence that supports modality specific representations. See also Prinz (2002) for a philosophical defense of concept empiricism.

perception and cognition share common neural systems, they function simultaneously in the same mechanisms and cannot be divorced" (Barsalou, 1999, p. 603).

According to the view of perceptual symbol systems the role of perception in the development and use of conceptual knowledge is fundamental. In relation to the 'dog' example mentioned previously, the distinction between perceptual qualities like 'fur,' barks,' 'has four legs,' etc. and presumably non-perceptual functional qualities, like making a good companion, does not hold. Although 'making a good companion' is more complex and not easily identifiable with any simple perceptual quality, it does not rule out a perceptual basis for that kind of conceptual knowledge. The quality of making a good companion can be associated with a number of perceptually derived situations. For example, a good companion gets fed, brushed, taken for walks, etc. These are perceptual qualities of good companionship.

It is important to point out that there is no claim in Barsalou's theory that there must be a simple mapping function that relates an item of conceptual knowledge to a specific perceptual representation in a given cortical area in the brain. A claim of this nature would tend to have a rather static and rigid view of the representation of conceptual knowledge. More specifically, such a static view maintains that conceptual knowledge exists as a coherent 'information package' ready to be activated by a certain stimulus. In this sense, activation of conceptual knowledge activates all information associated with the concept. Another implication of the view that concepts represent packets of organized information in long-term memory is that conceptual knowledge would be less amenable to contextual factors.

Perception according to Barsalou can be understood as providing a basis from which to categorize and think about aspects of our environment. But this should not be interpreted as a unidirectional relationship between perception and categorization, i.e., the view that perception feeds the process of categorization with the 'primitive' building blocks from which to reason about aspects of the environment. Schyns (1997) referred to this view as a *fixed feature view of categorization*. Schyns presents evidence to question this view of categorization showing that even perceptual organization is influenced by previous categorization experience. This means that if subjects are first given experience in categorizing visual stimuli according to predetermined instructions, they will on later tests of categorization tend to miss clear perceptual changes in the stimuli and instead process the stimuli in a way that is consistent with previous categorization training. Previous categorization experience affects the way we perceive objects in our surroundings. On the basis of such evidence, Schyns argues for a bi-directional influence of perception and categorization. Perception plays a role in categorization and categorization experience can influence the way we see things. His further point is that categorization is highly dependent on the nature of the task demands involved categorization.[11]

Perceptual symbol systems and the corresponding view of modality-specific representations maintain that categorization is inherently dynamic. Barsalou et al. (2003) actually go so far as to question the validity of concepts as a scientific construct. Their view is that the activation of conceptual knowledge depends critically on the situation and the task. If concepts exist, they exist as structured information in working memory with the purpose of solving a given task in a given situation. As a result, modality specific representations do not represent concepts; they represent conceptual knowledge in long-term memory that can be retrieved to produce behavior that appears as if we *possess* concepts as structures in long-term memory. In this sense then, concepts serve as the 'glue' that allows us act appropriately given a certain context.

I should point out that I have discussed two different ways of viewing concepts as the 'glue' that structures our mental world. The first has to do with the fact that concepts serve an important role in thinking. We use conceptual knowledge to make inferences, generalize and communicate. Conceptual knowledge is used to guide our impressions (understanding) and it is also used productively in order to affect changes in our environment. On the basis of our understanding of a given situation, we act accordingly. We can tune our actions to produce intended social and physical effects. This interpretation of concepts as 'glue' or 'timber' is what the authors of the introductory quotes intended. But given the fact that context and task constrain our use of conceptual knowledge, we still need to know to what extent our access to information that supports conceptual knowledge is constrained by perceptual processes, i.e., our information gathering resources. It is the interaction between these two aspects (the situational determinants of our conceptual abilities and perception) that constitutes the basis from which to understand the acquisition and use of conceptual knowledge. While perception and contextual factors could be viewed as the glue that holds concepts (conceptual knowledge) together, conceptual knowledge can be understood as the glue that holds thinking together. The contextual factors referred to here will be discussed in greater detail in the following sections.

2.2 Hierarchies, the Basic Level and Graded Structure

Three major phenomena have characterized a large portion of categorization research since the early 1970s. These are: the hierarchical structure of categories, the occurrence of a basic level of categorization and the graded structure of categories around a central instance, or prototype. In Chapter 5, each of these phenomena will be

---

[11] See also Goldstone (1994) or Goldstone and Barsalou (1998) for further discussions of this relation between perception and conception.

discussed in greater detail, and I will also relate them to the categorization of actions. Briefly, the hierarchical structure of categories refers to the taxonomical organization of 'things'. For example, my dog 'Zorro' is a Jack Russell terrier, which is a dog which is a mammal which in turn is an animal. In this case, the example goes from the specific to the general. The hierarchy represents class inclusion relations between the different levels in the hierarchy. Knowing that Zorro is a Jack Russell can provide information about the appearance and temperament of Zorro if one is familiar with Jack Russell terriers. And knowing that Zorro is a dog activates information about the dog-like qualities mentioned previously in the chapter. Zorro inherits the properties of the categories at more general levels of description.

The basic level refers to a privileged level in a taxonomic hierarchy. Results supporting the notion of a basic level show that people tend to identify, communicate and interact with objects on this level of abstraction (Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976; Rosch, 1978). According to Rosch (1978), the basic level represents a level at which different categories "mirror the correlational structure of the environment" (p. 31). In other words, basic level categories reflect natural divisions between kinds of objects we experience in our surroundings.

Regarding the notion of graded structure, some instances within a given category seem to be more typical or more representative than other instances (Rosch, 1975; Mervis, Catlin & Rosch, 1976). For example a chair is quite typical as an instance of furniture whereas a piano is less typical. The finding that different category instances, or exemplars, are more or less typical of a given category indicates that there is a 'typicality gradient' for that category. Another way of describing the phenomenon is to say that the category exhibits graded structure. The graded structure of categories has been investigated using a number of different methods, some of which will be described below. Related to the graded structure of categories is the finding that typicality is graded with respect to a central or prototypical instance of the category. A prototypical instance does not necessarily have to be a specific concrete instance that has been previously encountered. A prototype can be a kind of combination of shared features between members of a category. In this sense prototypes can be abstract; they do not have to be specific concrete instances that are naturally found. A prototypical chair for example consists of shared features from chair instances that serve to relate all instances of chairs within the same category.

2.3 Conceptual Knowledge of Actions

Given the ecological importance of being able to categorize the actions of others and the apparent perceptual salience of motion based actions, the question arises as to how action categories are structured in regard to the three above mentioned phenomena. There is very little research that specifically investigates the nature of action categories according to the discussed phenomena. As mentioned previously, results supporting the hierarchical structure of categories, the basic level and graded structure and prototypes have been obtained by using static images of naturally occurring objects, human artifacts and artificially constructed stimuli. Why should the nature of action categories be any different from these categories? This question should not be taken to assert that there may be a single fundamental difference between actions and objects that leads to differences for all three categorization phenomena. It is possible that differences between actions and objects are not relevant to the categorization phenomena. For example, the inherent spatiotemporal features of motor based actions may not lead to fundamentally different ways of organizing conceptual knowledge of actions. The upshot of this is that, although the characteristic features of objects and actions differ, this does not necessarily imply that it will lead to drastically different results regarding the categorical organization of actions.

It is quite possible that categorical organization for vastly different domains is governed by very similar principles. For example, Vigliocco, Vinson, Lewis and Garrett (2004) claim that the semantic representation of nouns and action verbs can be described according to a common measure of semantic distance that shows comparable effects of similarity. The results from testing their model ‘Featural and Unitary Semantic Space’ (FUSS) show that despite having different features that can lead to differences in the hierarchical structure of nouns and verbs, the lexico-semantic space of objects and actions can be modeled according to the same principles (cf. Gärdenfors, 2000, 2007). The important issue here concerns the extent to which action categories exhibit a hierarchical structure according to a taxonomical organization vs. another more feature based or context based organization.

Before addressing the categorical organization of actions, we need to first take look at the features of actions that might be used to categorize them. According to Tranel, Kemmerer, Adolphs, Damasio and Damasio (2003), action concepts include knowledge about the behavior of entities; people and animals as well as artifacts (tools) that are used by humans to achieve goals. They further describe conceptual knowledge about actions in terms of basic components, or dimensions. These include the following: *causal organization* (transitive versus intransitive movements, e.g., hit vs. arrive, go or fall), *body-internal behavior* (running, walking, etc.), *change of state* in the location or state of another object through direct contact (lifting or ironing), *the use of specific body parts* in an action (waving, throwing, speaking, grasping, etc.), *spatial trajectory* (different spatial patterns associated with the manner in which an action is performed, e.g., jogging vs. sprinting), *temporal aspects* associated with different actions (throwing vs. stirring or waving) and the *goal of an action* (running may serve different goals like chasing something, fleeing or getting some exercise). A final component has to do with the *emotional content* of an action. Actions can also convey emotions.

While the components identified by Tranel et al. (2003) appear to capture many relevant aspects of actions, some components appear to play a greater role in our understanding of actions and words that denote actions or interactions with objects.

For example, the use of a specific body part differentiates leg-based actions (running and kicking) from hand-based actions (waving and throwing). Mouth- and face-related actions are also distinguished from other actions on the basis of the body-part component. The body-part component appears to be neurologically instantiated in the somatotopic organization of the motor area (Pulvermüller, 2001). Activation of motor neurons for leg and foot movements occurs in an area anatomically superior to face and arm movements, while activation of arm movements is anatomically superior to neurons for face and mouth movements. This suggests that action categories differentiated on the basis of body parts might differentially activate the different motor areas corresponding to the different body parts. Using behavioral measures and EEG recordings, Pulvermüller, Härle and Hummel (2001) obtained results consistent with this hypothesis. In their experiments, subjects were presented with action words (verbs) depicting leg-, arm- and face-related actions. Intermixed with the action words, subjects also saw pronounceable pseudowords. Subjects were given a lexical decision task where they simply pressed a button in response to a word, but not to the pseudowords. The behavioral data showed that subjects responded faster to the face-related words than the leg-related words. Pulvermüller et al. explain the difference in response times as being due to the different neurological organization of verbs referring to leg-, arm- and face related actions. The wider cortical distribution of neurons for leg-related actions leads to longer response times. On the basis of the EEG recordings and subsequent analyses, they also found activation for leg- and face-related actions in respective areas along the motor cortex. These results indicate a neurological basis for broad categorical distinctions between leg- and face-related action categories. The evidence for arm-related actions was confounded by the fact that subjects were using their arms in the response. This prevented Pulvermüller et al. from drawing conclusions about the activation of arm related areas in the motor cortex.

The distinction between leg-, arm- and face-related movements was also investigated in an fMRI study by Buccino, et al. (2001). They let subjects view video sequences of biting an apple, grasping a cup or an apple and kicking a ball or pushing a brake. Subjects also viewed similar actions that did not involve an object. The results showed that viewing both object and non-object related actions led to significant somatotopic activation in premotor areas. Leg-, arm- and face/mouth-related actions activated distinctly different areas. Buccino et al. also found a somewhat different pattern of activation for object and non-object related actions. While non-object related actions also led to somatotopic activation, inferior parietal areas were somatotopically activated as well for the object-related actions. The activation of inferior parietal areas indicates an action-object function coupling. On this level, the function of an object is tied to the potential for interacting with an object.

2.4 Concept Hierarchies

In order to develop an understanding about the organization of action categories, it may be the case that action categories share a similar structure with object categories. Given the wealth of research on the categorization of static objects, I will discuss a number of findings that seem to be relevant for action categories.

2.4.1 Object Domains

One important aspect of the organization of conceptual knowledge has to do with the hierarchical structure of concepts. Natural kinds and artifacts can be classified according to different levels of generality or inclusiveness, i.e., a taxonomic organization as mentioned previously. This hierarchical structure can be interpreted as facilitating our thinking about objects and entities. The facilitation arises out of the relations between levels of the hierarchy and the information associated with the different levels. Knowing what a mammal is and that a bat is a mammal, allows us to distinguish it from birds and group it together with other animals that nurse their young. It should be emphasized that the hierarchical nature of concepts referred to here is not necessarily the same as a scientifically based hierarchical system of classification like the Linnean system. The hierarchical structure discussed here is more of a folk psychological classification scheme indicating the common knowledge that people have about natural kind and artifact taxonomies.

At the general level of classification, natural kinds can be classified as living things. Living things can in turn be divided into categories of plants and animals, which can in turn be divided into further subcategories, e.g, mammal, dog, terrier, etc. This is illustrated in Figure 2.1. The hierarchical organization of concepts is indicated by the vertical dimension in the figure. One way of viewing the relation between levels in the hierarchy is in terms of *superordinate* (or hypernym) and *subordinate* (or hyponym) levels. Levels higher up in the hierarchy are superordinate, and lower levels are subordinate in relation to higher levels. In addition to this basic distinction, hierarchies often include a basic level, a level of abstraction between general superordinate categories and very specific subordinate categories.

A concept hierarchy can be viewed as a network of conceptual knowledge in long-term memory. The nodes in the hierarchy represent conceptual knowledge associated with a specific concept on that level. According to this view, not all conceptual knowledge associated with a specific concept is stored on that level. A subordinate level concept inherits the properties from superordinate level concepts. The additional information stored at lower levels distinguishes it from superordinate level concepts. For example, the concept of DOG contains conceptual knowledge about dogs that distinguish it from contrast categories of other mammals, and the conceptual knowledge that is associated with all mammals is inherited via the inclusion relation between MAMMAL and DOG. In this sense, the links between nodes represent IS-A relations between levels in the hierarchy such that activation of a subordinate level node can in turn activate information on higher levels when presented with a task that requires hierarchical processing (Collins & Quillian, 1969).

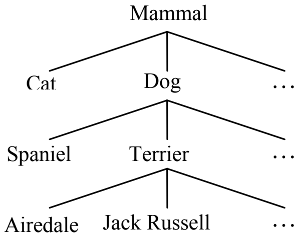


Figure 2.1. A category hierarchy illustrating vertical and horizontal dimensions.

In contrast to storing concept hierarchies, another way of viewing the hierarchical nature of concepts is to maintain that conceptual knowledge consists of feature lists that are used to make categorical inferences. This view asserts that hierarchical relations are not pre-stored in long-term memory but rather determined by accessing the features associated with concepts in the context of a task that requires knowledge of concept hierarchies. Specifically, the concept of DOG then includes, rather than inheriting, the superordinate level features of gives birth to living young, nurses its young, breathes, etc. (Murphy, 2002). The reason for mentioning these two views is to show that the issue of the representation and processing involved in the hierarchical nature of concepts is not resolved. Consequently, if the issue is not resolved for object domains, then it is quite likely to be equally problematic for other conceptual domains as well. This should not be understood as casting doubt on the hierarchical nature of conceptual knowledge in general. The issue is rather one of how knowledge of concept hierarchies is structured in long-term memory.

The mention of feature lists (or attributes) associated with concepts on different hierarchical levels raises the question of the extent to which feature lists can be used to differentiate between different levels in a concept hierarchy. In their influential study of object categorization, Rosch, Mervis et al. (1976) instructed subjects to list the attributes of a limited number of biological and nonbiological objects (experiment 1). When presented with an object name, subjects wrote down as many attributes as they could think of. The object names corresponded to different levels in a category hierarchy. For example, *tool* was presented as a superordinate level category whereas *hammer* was presented as a middle or basic level category, and *claw hammer* was a subordinate level category. In contrast to *tool*, an example from a biological category was *tree* (superordinate), *maple* (basic level) and *sugar maple* (subordinate).

The results showed that the number of listed attributes (or features) varied as function of category level. Subjects listed significantly fewer attributes for objects named at the superordinate level than for lower levels in the hierarchy. Furthermore, the largest increase in listed attributes occurred between the superordinate and the basic level. There was no significant increase in listed attributes between the basic and subordinate levels. It should be noted however that this finding only occurred for the nonbiological categories. Attribute lists for the proposed superordinate level for biological categories showed that subjects listed relatively many attributes such that there was no reliable difference between the hierarchical levels. Rosch, Merivs et al. (1976) interpreted this to mean that the proposed superordinate level (*tree*, *bird*, and *fish*) were more like basic level categories.

A further aspect of the list of attributes suggests that the frequency of different kinds of attributes varies according to the different hierarchical levels. For superordinate level categories, Rosch, Mervis et al. found a majority of 'functional' attributes. Although they do not describe the functional attributes, it is not difficult to imagine what they might have been. Functional attributes for the category of *tools* might reasonably include 'used for fixing things,' or 'used for building things.' In contrast to the superordinate level, attribute lists for basic level objects tended to contain more nouns and adjectives. Significantly more nouns and adjectives were produced for basic level than for superordinate level categories. For subordinate level categories, adjectives occurred more frequently than for basic level categories, which indicates that it is the modification of features that distinguish objects on the basic level from objects on a subordinate level.

Differences between levels in a concept hierarchy have also been demonstrated using processing times to verify the membership of nonbiological objects on different hierarchical levels. For example, Rosch, Mervis et al. (1976) presented subjects with category labels on superordinate, basic and subordinate levels, e.g., *furniture*, *table*, *kitchen table* respectively. Shortly after hearing a category label, subjects then viewed a color photograph of an object. The task was then to indicate as quickly and accurately as possible whether or not the depicted object belonged to the previously heard category. The results showed that verification response times were fastest when a picture was preceded by a basic level category label, e.g., *table*. When presented with a subordinate level label, subjects' verification times were significantly slower than for both basic level and superordinate level labels. The differences in verification times indicate that the hierarchical structure of categories affects processing times in categorization tasks. The findings also point to a processing advantage for the basic level.

A further aspect of the basic level has to do with linguistic output. People tend to use words on the basic level to name objects. In an object naming task, Rosch, Mervis et al. (1976) let subjects view pictures of objects (biological and nonbiological). Three contrast sets were constructed. In the superordinate contrast set, subjects saw 9 examples from 9 different superordinate level categories. For this set, a superordinate level label would be sufficient to distinguish between the pictured objects. The basic level set consisted of one picture from each basic level category. In this case, a basic level label would be enough to uniquely identify each object from the others. Similarly, for the subordinate set, subordinate level labels were needed in order to distinguish the objects from one another. Despite the fact that different labels were sufficient to distinguish the objects from one another, subjects clearly preferred to name the objects using basic level labels. Rosch, Mervis et al. were also able to rule out the effect of the frequency of the words as an alternative explanation to the results. Furthermore, Rosch, Mervis et al. (1976) tested whether the obtained results could be explained by the fact that subjects did not know the correct superordinate or subordinate names of the objects. The results from this investigation revealed that subjects were able to confirm the superordinate and subordinate labels for the objects. This shows that the preference of basic level labels was not merely a consequence of not knowing the superordinate or subordinate level names of the objects. Similar results regarding linguistic output were found for the naming of events (Morris & Murphy, 1990).

2.4.2 Flexibility of Hierarchical Levels

Different hierarchical levels suggest that people organize their conceptual knowledge according to various levels of abstraction. This, however, should not be interpreted to mean there is a single hierarchy consistently obtained and used in all contexts and situations. As discussed above, there is evidence to suggest that objects can be classified differently depending on the context and specific tasks. There may be multiple hierarchies that reflect flexibility in the psychological construction and use of levels in a conceptual hierarchy.

In an attempt to replicate the findings from Rosch et al. (1976), Murphy and Brownell (1985) also let subjects categorize visually presented nonbiological objects in relation to different levels of category labels. In addition to the different hierarchical levels, Murphy and Brownell also manipulated the typicality of the objects by choosing highly typical objects (e.g., desk chair) and atypical objects (e.g., beach chair). The purpose of this manipulation was to investigate the extent to which category verification is influenced by our ability to differentiate between objects on the same hierarchical level. The reasoning behind their experiment was that as objects become more atypical, they also become more distinct or differentiated from the more typical objects in the category. This increase in distinctiveness should allow subjects to make quicker verifications for subordinate level objects. When the objects were typical for the category, Murphy and Brownell found results similar to Rosch, Mervis et al. For atypical objects, however, category verification was faster on the subordinate level than for the basic level. Murphy and Brownell suggest that "… giving a taxonomic level a single label, such as *basic* or *subordinate* is too simplistic and that there is a continuum of category "basicness," as Rosch et al. (1976) originally speculated" (p. 73, italics in original).

The flexibility of hierarchical levels can also be demonstrated in experiments that study the effects of different levels of domain expertise. For example, in 3 experiments, Tanaka and Taylor (1991) showed that experts within a given object domain, e.g. avid birdwatchers, tend to organize their knowledge at a subordinate level rather than at a basic level. In the first experiment, bird and dog experts produced feature lists in response to viewing superordinate, basic level and subordinate level category labels (e.g., *animal*, *bird*, *robin*). Both groups of experts produced feature lists for bird and dog domains. While dog experts had domain specific knowledge about dogs, they did not have expert knowledge of birds, and vice versa. When the subjects were presented with category labels from their non-expert (novice) domain, basic level labels elicited more features than subordinate level labels. Within their domain of expertise, however, the subjects listed significantly more features for subordinate level labels than they did for the domain in which they were non-experts.

Tanaka and Taylor (1991) also found that experts were more inclined to produce subordinate level names in response to pictures in a free naming task when the pictures depicted objects from their respective areas of expertise. In a third experiment, the experts were given a speeded category verification task, similar to the one used in Rosch, Mervis et al. (1976). For this task, the experts had to verify whether a depicted animal belonged to a previously presented category (superordinate, basic level or subordinate). When the depicted animal came from their novice domain, verification times were fastest for the basic level category labels. When, however, the depicted animal came from their domain of expertise, verification times were equally fast for subordinate and basic level category verification.

According to Murphy (2002) experts have likely developed a greater sensitivity to underlying differences between subordinate level objects. This increase in sensitivity leads to greater differentiation for objects categorized at the subordinate level. Murphy (2002) points out that the effect of expertise on categorization should not be viewed as a shift in the basic level from for example *bird* to *robin* but rather that expertise involves an increase in categorization flexibility. Bird experts can just as easily use the basic level as they can subordinate level categories. Johnson and Mervis (1997) also present findings of a basic level shift.

An alternative to a single taxonomic organization of objects is the organization of objects in terms of script categories. Script categories reflect the structure of routines used in different situations. A further aspect of script categories is the relation to a specific goal. For example, although ski pants are an article of clothing, they are also associated with the category of winter clothing and more specifically with clothing to wear when skiing. This is an example of the cross-classification of objects. Classification is not restricted to a single hierarchy, but can be influenced by specific situational goals. Conceptual knowledge about objects can be used to appropriately classify objects according to daily routines. (See also Barsalou (1982, 1983) for related findings.)

In an extensive investigation of the extent to which the organization of categories for different kinds of food items are restricted to a single taxonomy, Ross and Murphy (1999) found a clear tendency for subjects to generate script categories about as often as taxonomic categories. According to a taxonomic categorization, eggs are an instance of a dairy product. However, on the basis of a script category, eggs are often viewed as an example of a breakfast food. Ross and Murphy found that script categories were a reliable basis on which to make inferences about different kinds of food. The findings of cross-classification suggest that taxonomic organization is not the only way we organize conceptual knowledge about objects in our environment.

In summary, there is clear evidence for a hierarchical organization for object categories according to taxonomic levels. There is also clear evidence that a single taxonomic organization does not fully account for the categorization behavior of people. If this is the case for object categories, to what extent might action categories exhibit a taxonomic organization?

2.5 The Hierarchical Structure of Action Categories

What evidence is there to suggest a hierarchical structure for action categories? How are action categories organized psychologically? Is there one structure that captures the folk psychological organization of action categories? Or is the psychological organization of actions best described as inherently flexible, dependent upon the context and situational demands and tasks? Is it more like the cross-classification in the food example mentioned above?

Intuitively, it seems that action categories might be hierarchically structured in a way similar to object categories. For example, the general category of *locomotion* includes walking, running and crawling. And *walking* can in turn include different ways of walking like strolling, marching, plodding, staggering, limping, etc. Ingesting can be done by drinking and eating. Sipping and gulping are instances of *drinking*. Given the paucity of research dealing specifically with the hierarchical structure of action categories, it is unclear to what extent these examples actually mirror the cognitive representation of action categories. Some insight however can be gained from looking at findings from three related areas: differences between the cognitive organization of hierarchies for nouns and verbs, the hierarchical structure of event concepts and findings of categorical effects of action naming using displays of biological motion.

2.5.1 Nouns and Verbs

Although differences between the cognitive organization of hierarchies for nouns and verbs can give us some insight into the structure of action categories, some degree of caution is required. The representation and processing of verbs is not the focus in this book. Verbs as lexicalized concepts for actions are likely to vary between languages, and making the assertion that lexicalized items directly map onto the underlying conceptual representations runs the risk of making what Bock (1996, as cited in Vigliocco et al., 2004) referred to as the “mind in the mouth” fallacy. In steering clear of this fallacy, it can still be the case that verb structure can provide some information about the underlying hierarchical structure of action categories. Where does verb meaning come from if not from being grounded in the conceptual representation of actions?

In a study comparing the semantic relatedness of nouns and the relatedness of verbs, Huttenlocher and Lui (1979) found that memory for the relatedness of nouns was greater than for the relatedness of verbs in adults and children. One explanation they give is due to the different hierarchical organization of nouns and verbs. Nouns they claim are organized into well structured hierarchies whereas the organization of verbs is more “matrix-like.” The matrix like organization referred to here is similar to the cross-classification results mentioned above. Elements of verb meaning cut across different semantic fields. This has also been shown in the results from naming norms for actions and objects where name agreement for pictures of actions was lower than for objects (Bonin, Boyer, Méot, Fayol & Droit, 2004). This suggests that the naming of actions is more variable than the naming of objects.

Vinson and Vigliocco (2002) also found differences between the clustering of semantic fields for actions and objects. While the distance between semantic fields for the object domain showed clear category boundaries, clusters for semantic fields for actions tended to be more evenly distributed across the representational space.

The results from these studies indicate that the underlying organization of action and object categories is somewhat different. This, however, does not rule out the idea that even action categories can be represented hierarchically, although perhaps not to the same extent as object categories. The hierarchical organization of action categories will be developed in greater detail in Chapter 5.

2.5.2 Taxonomic Hierarchies for Events

The domains of events and actions are similar in that they both have a temporal as well as a spatial dimension. Events are also context sensitive, being influenced by the goals of an actor (Zacks & Tversky, 2001). Given these similarities, evidence of a hierarchical taxonomic structure of event categories would suggest that action categories may too be structured in a similar way. In their studies on converging operations in event taxonomies, Morris and Murphy (1990) investigated the generation of feature lists, similarity ratings for events, event verification and event naming according to different taxonomic levels in event hierarchies. The results showed a clear effect of the different taxonomic levels. Two examples of the events and taxonomic levels are presented in Figure 2.2. In the feature listing task, subjects were presented with event labels at the different taxonomic levels and asked to list the actions common to all examples of the event. The results showed a significant increase in listed features from the superordinate to the middle level. This can be explained by the general nature of superordinate level events. It is relatively difficult to find features that are common to all examples of entertainment compared to movies. There was no increase in features from the middle to the subordinate level, indicating that subordinate level events do not significantly add to the information conveyed by features on the middle level of the hierarchy. This result is similar to the results for nonbiological object categories obtained by Rosch, Mervis et al. (1976).

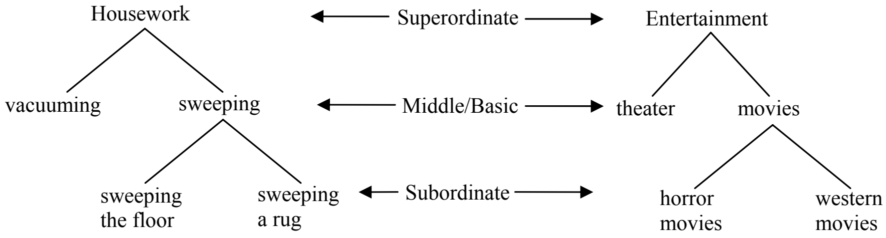


Figure 2.2. Examples of event hierarchies in Morris and Murphy (1990).

In their third experiment, Morris and Murphy also found that subjects were fastest at verifying whether an action was a part of a named event at the middle level compared to the superordinate level. For example subjects were fastest at verifying "scream during the scary parts" when the event label was 'movie' than when the event label was 'entertainment.' Similar to the feature listing results, there was no significant difference between the middle and subordinate levels. Morris and Murphy did however obtain a difference between the middle and subordinate levels in an event naming task. The purpose of this experiment was to see what labels subjects would use to name different events. Subjects were given short stories about events. After reading the events, they were instructed to simply name the events. The results showed that subjects clearly preferred middle level event names. Even when a subordinate level name was more appropriate in order to distinguish between subordinate level events, subjects produced slightly more middle level names. It appears then that there are important similarities between the taxonomic organization of events and objects. (See also Rifkin (1985) for similar results.)

2.5.3 Action Categories and Biological motion

The most direct assessment of the categorical effects of action perception has been done using displays of biological motion. Employing the point-light technique developed by Gunnar Johansson (1973, 1975) to study the perception of biological motion, Dittrich (1993) let subjects view a number of different actions depicted in the point-light displays and simply indicate when they recognized the actions. Dittrich investigated categorical effects by including actions from 3 different superordinate categories (*locomotion*, *social actions* and *instrumental actions*). Instances of *locomotion* were walking, going upstairs, leaping and jumping. The category of *social actions* included dancing, boxing, greeting and threatening. *Instrumental actions* included interactions with objects: hammering, lifting a box, bouncing a ball and stirring. Subjects first indicated recognition by pressing a button and then they provided a verbal label for the action. The results revealed differences in reaction times for the different categories. Locomotory actions were generally recognized more accurately and faster than the instrumental actions and social actions. The social actions were also recognized faster than the instrumental actions. Subjects had the most difficulty with the instrumental actions. In terms of the taxonomical organization of action categories, the results suggest differences between superordinate level categories. This finding does not specifically provide evidence for the hierarchical structure of action categories, but rather suggests that superordinate level distinctions between action categories play a role in action perception. (See also Dittrich (1999) for further discussion of this issue.) Even though category-level distinctions may be less clear for actions than for objects, they are sufficiently clear to be an important factor in our perception and conceptual organization of actions. This theme will be addressed in the empirical studies presented in this book.

The evidence from the noun-verb studies suggest that the action categories referred to by verbs are not as clearly hierarchically organized as nouns that refer to objects. Unfortunately the evidence only speaks to the *relative* hierarchical organization of nouns and verbs and does not address the *extent* to which verbs are actually hierarchically organized. In one sense, however, a more “matrix-like” organization seems reasonable given the service of actions in achieving different goals. The evidence from event categories appears to support the idea that event categories have a hierarchical organization, similar to the way object categories are organized.

Let me address the specific questions posed at the beginning of section 2.5. The results from studies on event categorization suggest that even action categories are hierarchically organized. If events, which are even more dependent on a specific goal, exhibit a hierarchical organization, then the natural actions being addressed here will likely also exhibit a hierarchical organization. However, it is not likely the case that there is only one hierarchical organization that captures the psychological representation of action categories.

2.6 The Basic Level

In this section, I will first discuss the basic level from the perspective of the visual form of objects and the role that parts seem to play in the visual form of objects and our interaction with them. [12]

The findings of Rosch, Mervis et al. (1976) supporting a taxonomic organization of object concepts (see section 2.4.1) according to hierarchical levels also point towards the basic level as a privileged, fundamental level of categorization. One important aspect of this privileged status is the role of visual perception. To the extent that perception is constrained by the physical structure of the objects within a physical environment, categorization, and particularly the basic level, may be correlated with the structure of the environment. In contrast to superordinate and subordinate level categories, basic level categories are claimed to be most differentiated in that the members of basic level categories have many features in common but have relatively less features in common with the members of contrasting categories. In addition to the results from attribute listing and category verification tasks, Rosch, Mervis et al. found that subjects listed very similar motor movements for objects categorized at the basic level, suggesting that our interaction with objects is best understood at the level of *hammer* rather than *tool* or *claw hammer*. Objects categorized at the basic level also appear to be visually more similar in overall shape than objects at a superordinate level. There was, however, significantly greater image overlap for subordinate level categories compared to the basic level. This difference was much smaller than the difference between basic and superordinate levels.

In a further investigation of the role of object shape in the categorization of objects, Rosch, Mervis et al. (1976) let subjects view the average shape of objects on the superordinate, basic and subordinate levels. (See Rosch, Mervis et al. (1976) for details.) Subjects were provided a list of categories for each shape and instructed to circle the category to which they thought the shape belonged. Subjects were also instructed to write down their best guess of the depicted object. The results showed that subjects were no better than chance at identifying superordinate level objects. Object shapes constructed according to the basic level lead to significant identification for superordinate and basic level categories. For the object shapes based on the average of subordinate level objects, subjects were no better at identification than for the basic level objects. Despite the previously mentioned result of significantly greater shape overlap for subordinate level objects, it did not lead to better identification for subordinate level objects.

---

[12] Murphy (2002) makes an important point about terminology when he asserts that it is not the objects themselves that are basic level, but rather it is the categories that can be considered as superordinate, basic or subordinate. Concepts and categories are *psychological constructs*. Despite my usage of the phrase ‘basic level objects,’ I agree with Murphy. I will, however, sometimes use ‘basic level objects’ to refer to objects categorized at the basic level.

In two further experiments with an emphasis on the visual nature of object categorization, Rosch, Mervis et al. (1976) first presented (primed) subjects with a spoken category label on the superordinate, basic or subordinate level. The underlying assumption in these experiments was that the category label would activate a representation (image) of an object representing the category. If the activated representation facilitated a categorization response, then there would appear to be some correspondence between the activated representation and the stimulus to which a response is required. Since superordinate level category labels (*tool*, *clothing*, *vehicle*, etc.) pick out objects that are visually very different, it was predicted that there would be no facilitation when primed by a superordinate level category label. In contrast, basic and subordinate level categories refer to objects that are visually similar (*hammer*, *pants*, *car*). Consequently, Rosch, Mervis et al. predicted facilitation for basic and subordinate level category labels. In one experiment, subjects were presented with a card. They were told that an object would be presented randomly (either right or left) on one side of the card. An abstract drawing was presented on the other side. Prior to viewing the card, subjects heard a category label, and upon viewing the card, they were to indicate the side on which the object appeared. The findings were consistent with the prediction. Object detection was significantly better when it was preceded by a basic level category label compared to a superordinate level label. There was, however, no significant difference in detection facilitation for basic and subordinate level labels.

In their second priming study, Rosch, Mervis et al. (1976) predicted that a basic level category label would also facilitate determining whether or not two depicted objects were physically identical. Because a basic level label activates a visual representation of an object that can represent the entire category, it could be used to judge the physical similarity of two objects. The findings from this study mirror the findings from the previous study. There was no significant priming for superordinate level labels. Basic and subordinate level labels, however, led to significant priming, and there was no priming level difference between basic and subordinate level labels.

The results from these studies indicate that the visual shape of objects plays a significant role in categorization at the basic level. While visual shape is also an important factor on the subordinate level, it does not seem to lead to a processing advantage for subordinate level objects. It also appears that the visual shape of objects on the basic level can be captured by a mental image, which can facilitate detection and judgments of physical identity. The results from the previously mentioned verification study also indicate that visual verification of objects is fastest at the basic level. This, however, needs to be seen in the light of the results from Murphy and Brownell (1985) (see section 2.4.2) where they showed that verification times for atypical objects on the subordinate level were faster than basic level objects.

2.6.1 The Role of Object Parts in the Basic Level

A further important perceptual aspect of the basic level is the role that object parts play in the shape of objects and the role parts play in our potential interaction with objects (Tversky & Hemenway, 1984). Compared to the superordinate level, basic level objects tend to have well-defined parts. Furthermore, different basic level objects can be distinguished on the basis of their parts. Objects on the subordinate level also have well-defined parts, but they are shared by other subordinate level objects within the same basic level category. Subordinate level objects tend to differ more on the basis of the ways in which parts are differently modified rather than on the basis of altogether new parts. For example, a sports car and a sedan share the major parts of an automobile but they differ according to styling features that do not affect the general function of driving (although a sports car may offer a very different driving experience). Object parts then "play a special role in the vertical organization of categories, that of distinguishing the basic level of reference" (Tversky & Hemenway, 1984, p. 186).

The findings of Tversky and Hemenway (1984) also suggest a perceptual basis for determining the function of objects. In their studies, the perceptually salient parts of objects were related to the function of different objects. For example, the parts listed for TABLE included legs, top, surface and wood. For GUITAR, subjects listed strings, tuning keys, neck hole and wood. The exception in these cases of parts reflecting the functional properties of objects is the attribute of wood. 'Wood' as a functional part of tables and guitars is arguably less important than the other listed parts. These results fit nicely with the findings in Rosch, Mervis et al. (1976) regarding the similarity of motor movements associated with objects categorized at the basic level. The visual shape of basic level objects as determined by their visually salient parts and the role that the parts play in the potential for interacting with the objects support the notion that object function, rather than being an abstract property, is provided to an important extent by the partonomic organization of objects.[13]

2.6.2 Basic Level Actions

Results from previous research show that there is converging evidence for a basic level for objects. This converging evidence comes from different areas such as feature listing/similarity judgments, motor routines used in the interaction with objects, the visual form of objects, category membership judgments, word use and word structure (Murphy, 2002). It is not my contention to show here that there is a privileged level at which action categories exhibit all of these basic level effects. I do, however, present evidence that is consistent with a basic level for action categories. A further contribution regarding a basic level for action categories is to relate findings from the

---

[13] For a critical discussion of the role of parts in the categorization of objects on the basic level, see Murphy (1991a, 1991b) and the reply by Tversky and Hemenway (1991).

basic level for object categories with the purpose of generating research issues about the psychological organization of action categories.

In addition to recognizing and interacting with objects, much of our daily activity involves recognizing the actions of others and interacting with other people on the basis of our ability to see what they are doing. Is there any evidence to suggest a psychologically privileged level of categorization for natural action categories? In relation to the role of object parts mentioned above, is there any reason to suggest that actions may also be viewed as consisting of visually salient parts? My response to the first question is that there appear to be no empirical studies that specifically address this question. The previous references to Morris and Murphy (1990) and Rifkin (1985) indicate a privileged level of categorization for the broader notion of events but do not address the issue of such a level for natural actions. The research of Zacks and Tversky (2001) has also presented findings suggesting that events as well as objects categorized at the basic level are characterized by having good parts.

Despite the lack of research that addresses the notion of a basic level for categories of natural actions, there is some research that addresses our ability to make category judgments based on the spatiotemporal form of actions. In addition to this research, I will discuss findings that explore the role of body parts in our conceptual knowledge of actions. The purpose of this discussion is to show that the spatiotemporal form of actions as well as body parts appear to be important factors in the perception and categorization of human actions.

It is important here not to confuse the notion of a basic level in a taxonomic hierarchy for actions with the notion of a partonomic hierarchy for the human body. While body parts can be viewed in terms of a partonomic hierarchy and may play a role in the visual salience of potential basic level actions, it does not mean that the partonomic organization of the human body maps directly onto a taxonomic hierarchy.[14]

2.6.2.1 The Role of Body parts in Natural Action Categories

Regarding the role of body parts, one approach to action perception when viewing actions depicted in point-light displays is that visual processing proceeds in a hierarchical fashion (e.g., Johansson, 1973; Marr & Vaina, 1982; Webb & Aggarwal, 1982; Aggarwal & Cai, 1999). According to this approach, human body parts represent (semi-) rigid segments connected by the joints of the human body. These segments, once detected, are combined in a hierarchical manner to recover a figure that represents the human body. A consequence of this view is that action recognition *depends* on local processes involved in the detection of body parts, e.g., ankles, knees, hips, wrists, elbows, etc. In contrast to this approach, there is evidence to suggest that

---

[14] See Zacks and Tversky (2001) for a discussion of this issue for event structure.

coherent human motion can be detected despite impaired access to the local motion of body parts.

If local motion elements are necessary for the detection of a human body in motion, then preventing access to the local motion elements should severely impair detection of a point-light walker. Bertenthal and Pinto (1994) used a dynamic masking technique (Cutting, Moore & Morrison, 1988) to test this hypothesis. The gist of the masking technique is to copy the individual trajectories of each motion element (point-light) and then randomly superimpose the elements on the display together with the target object, i.e., the point-light walker. If the detection of a point-light walker depends on the extraction of local motion elements, then detection should be very difficult because there are multiple instances of the same local motion trajectories. Bertenthal and Pinto (1994) found that despite multiple copies of the motion trajectories in each display, subjects were reliably able to detect the presence of a point-light walker. Even when the motions of body limbs were masked, subjects could still detect the point-light walker. This indicates that the detection of body parts does not *precede* the perception of the global form of a point-light walker. It is important to point out, however, that Bertenthal and Pinto (1994) do not claim that the perception of body parts are not involved in the perception of biological motion. It is rather the case that they are not necessary.

In a further investigation of biological motion perception using the dynamic masking technique, Pinto and Shiffrar (1999) showed that, although not necessary, body parts may be sufficient for detecting the coherent figure of a point-light walker. In a series of experiments, they showed that subjects were able to detect the different subconfigurations of the human body even when masked. For example, when the target display only consisted only of the contralateral limbs (legs or arms), subjects were reliably accurate at detecting the presence of a coherent figure when it was embedded in a dynamic mask. Although detection of the subconfigurations was reliably better than chance, detection for some of the configurations was diminished in relation to detection for the whole point-light walker. Pinto and Shiffrar (1999) suggest that this finding be viewed in terms of the varying representativeness of the subconfigurations within the category of *human locomotion*[15]. Different subconfigurations, i.e., parts, of the human body, may be more or less representative of the human body as it engages in the action of walking. It is important to note here that although body parts may be sufficient for recognition, some form of hierarchical structure relating body parts to a whole appears to be necessary (Heptulla-Chattejee, Freyd & Shiffrar, 1996).

---

[15] Pinto and Shiffrar (1999) refer to the category as *human locomotion*. Since they only included a point-light walker in their study, I suggest that the subconfigurations should be viewed as exemplars of *walking*.

The results from Bertenthal and Pinto (1994) and Pinto and Shiffrar (1999) show that although local motion elements that correspond to the limbs of the human body are not necessary for detection of human form, they may be sufficient if some information about the hierarchical structure of the limbs in relation to the whole is available. The reason that they may be sufficient is that even the subconfigurations gave rise to successful detection. The reasoning here is that the subconfigurations of body parts can trigger the representation of dynamic human form. This seems to make sense in that we often successfully recognize the actions of others even when body parts are occluded or when we see the actions of others from different points of view. The subconfigurations of the human body may provide enough information to activate categorical information about the form of the acting object and about the specific action being performed. Another way of putting it is that there is sufficient information to provide access to an action prototype and thereby generate a sufficient match to the ongoing or previously carried out action. A discussion of action prototypes and the graded structure of action categories will be presented in section 2.7.

Human body parts may play a role in the activation of categorical knowledge of actions and thereby also provide a basis from which to distinguish between different natural actions. Recall for example the previously mentioned findings of Pulvermüller et al. (2001) and Buccino et al. (2001) showing differences in cortical activation for leg-, arm- and face-related actions. Their results suggest that the movement of specific body parts could indicate categorical breaks for actions. The question in the context of basic level categories is whether or not the role of body parts is distinctive for basic level action categories. I think there is reason to question that body parts per se can be used to determine a basic level for action categories. Intuitively, the role of body parts as constituting a basis for basic level action categories seems problematic because it would mean that very different arm actions (waving, throwing, saluting, stirring, sweeping, painting, shaking hands, bouncing a ball, etc.) would be considered subordinates to the basic level category of *arm-related actions*. The problem here is that a previous finding for object categories that distinguishes the basic level from the subordinate level is the relative distinctiveness between basic level and subordinate level categories. Basic level objects are maximally more distinct from one another than subordinate level objects (e.g., Mervis & Rosch, 1981). Although the categories of *arm- and leg-related actions* are quite distinct from one another, throwing and clapping as instances of arm-related actions also appear to be visually distinct from one another. On the other hand, different ways of *throwing* (e.g., lob, hurl, fling, flip, etc.) seem to be visually much more similar to one another than waving, throwing, stirring, clapping, punching, etc. The critical point here is that this gain in visual similarity among subordinates appears to be greater than the potential loss in visual distinctiveness in the move from *arm-related actions* to, e.g., *throwing* as a basic level action category. This thought experiment suggests that different arm-related actions (*waving*, *throwing*, *saluting*, etc.) might constitute basic level actions, and different *manners* of *throwing*, e.g., clapping, waving, punching, etc. constitute subordinate level actions.

A further objection to the idea that the motion of body parts constitutes a basis for basic level actions is the lack of linguistic output. We do not appear to communicate our actions or the actions of others by a lexicalized concept of ARM-RELATED ACTIONS. This factor of linguistic output and its indication of basic level categories will be discussed later on in this section.

In summary, there is some evidence suggesting that the parts of actions can be determined by specific body parts and the variations of their spatial trajectories during an action sequence.[16] The further question in regard to the role of body parts in determining a privileged level of categorization for actions is difficult to assess on the basis of the above evidence. The question here is to what extent do variations of the spatial trajectories of body parts contribute to determining a basic level for action categories. The effects of the similarity of actions that involve the same body parts, e.g., throwing, waving, will be a topic of the experiments in Chapter 5.

The perceptual features of objects and actions in terms of static and dynamic visual form as well as the role of parts of objects and actions appear to be important factors in our ability to categorize objects and actions. The perception of visual form and parts also appears to be important in the formation of categories at a privileged level of organization in concept hierarchies. The role of perceptual features is tied to the further notion that the perceptual features reflect structure in the environment. According to this reasoning, the perceptual basis of the basic level reflects a correlation between the structure of the environment and our ability to perceive that structure. An implication of this view is that we should expect basic level categories to be fairly similar across different cultures.

Is there evidence to suggest that basic level categories exhibit a high degree of cross-cultural stability? Malt (1995) reviewed a number of anthropological studies in order to answer this question. The cross-cultural studies she reviewed addressed the categorical coherence of biological objects, i.e., plants and animals. She found that very different cultures all seem to "consistently describe the smallest categories labeled with a primary lexeme as corresponding roughly to scientific genera" (p. 126). Malt (1995) also found that "cultures have either no terms above the folk generic level or a relatively small number" (p.126). "Similarly, no cultures are reported to have vocabularies of subgeneric terms near in size to those of generic terms; this mean that there are no cultures having a large vocabulary of terms corresponding to scientific categories below the level of the genus" (p. 126). In conclusion, Malt asserts, "The cross-cultural evidence on the existence of one primary, most salient level of

---

[16] See Casile and Giese (2005) for a computational model and psychophysical results that support the role of detection of motion discontinuities in action recognition.

classification given by the environment suggests that there is, in fact, one level that tends to be most salient cross-culturally" (p. 128).

This conclusion, however, should not be understood to mean that there is no room for individual variations in categorical structure within a given culture. To the extent that the knowledge of given categorical domain varies between individuals, it is quite likely that categorical partitions will reflect different levels of knowledge (Rosch, Mervis et al., 1976).

The finding of Malt (1995) regarding the level of genus as being most salient is consistent with the findings of Rosch, Mervis et al. (1976) where they originally hypothesized that *maple* and *oak* represent basic level objects but found instead that subjects' attribute lists indicated that *tree*, *fish*, and *bird* were more appropriately basic level. The issue has to do with the extent to which the most salient categorical level is the same level across different cultures. This is much stronger than simply saying that there is a most salient level of categorization, which is different for different cultures. That fact that all cultures may exhibit a most salient level of categorization is not evidence of cross-cultural stability for a preferred level of categorization.

The issue of cross-cultural stability for action categories will be addressed in Chapter 3. If the categorization of human actions is importantly determined by the constraints governing human movement and perception is sensitive to those constraints, then it may be the case that action categories will also exhibit a notable degree of cross-cultural stability.

2.6.3 Questioning the Status of the Basic Level

The notion that the basic level represents a privileged level of categorization has been vigorously challenged by Mandler (2004). She rejects the claims made by Rosch, Mervis et al. (1976) and others (e.g., Rosch & Mervis, 1975; Mervis & Crisafi, 1982) that their findings support the interpretation that basic level categories constitute a fundamental conceptual organization of knowledge about objects. Mandler (2004) suggests instead that a more appropriate interpretation of the results is related to what she calls perceptual categorization. Mandler's argument is based on the distinction between conceptual and perceptual categories. While conceptual categories involve complex conceptual knowledge and require conscious access to different kinds of information accumulated over time and experience, perceptual categories are formed "beyond the bounds of consciousness" (p. 291). As Mandler (2004) puts it:

> [Conceptual categories] are concerned with setting up kinds, that is, with formulating the sorts of things that dogs or tables are. Forming a concept is not automatic but rather is a focused and limited process. It appears to be serial in nature, with new information being added bit by bit, rather than accumulating simultaneously. (p. 292)

The perceptual processing involved in the formation of perceptual categories is automatic and operates in parallel. The previous findings of Rosch et al. (1976) relate to a level of perceptual organization rather than the conceptual understanding

(Mandler, 2004). Mandler maintains, for example, that infants are sensitive to perceptual dimensions (size and overall shape differences). Infants develop a perceptual schema by which they automatically process the physical dimensions of stimuli and use this information to distinguish between different objects. This sensorimotor ability is used to identify objects. But it is not the same as developing a conceptual understanding of the objects, which requires conscious access to knowledge about the kinds of things different objects are, e.g., an "information core" about knowing what a thing can do or what one can do to/with it within a given context (Mandler, 2004).

Mandler's (2004) view is that the basic level effects of Rosch, Mervis et al. (1976) are limited to the identification procedures (Smith & Medin, 1981) involved in object categorization. Basic level effects are not found for categorization when conceptual understanding (access to core meaning) is required to complete a categorization task.

A further difficulty with the notion of the basic level comes from VanRullen and Thorpe (2001) who showed that target detection for superordinate level categorization is much faster than would be expected on the basis of the previous findings of Rosch, Mervis et al. (1976). Subjects in their experiment were given pictures from two superordinate level categories (animals and means of transportation). Pictures of animals included different mammals as well as birds, fish, insects and reptiles. Means of transportation included pictures of cars, trains, trucks, civil and military aircraft, boats, hot-air balloons and rockets. The pictured objects were presented in natural scenes. Subjects had to simply indicate whether the scene contained an instance of the target category (animal or means of transportation). The results revealed remarkably fast reaction times. VanRullen and Thorpe (2001) reached the conclusion that if the delays in the motor pathways are taken into account, the visual processing involved in the successful completion of the task took around 150ms. In the context of the differences in processing between basic and superordinate level categories mentioned previously, VanRullen and Thorpe assert (p. 666), "The surprisingly good performance and very short reaction times obtained here cast doubt on the intuitive idea that visual processing would require a basic level identification of the stimulus before its potential superordinate level categorization (Rosch, Mervis et al 1976)."

What is the function of the basic level? Why is there a level of categorization that is more salient with regard to, for example, perception (visual form) and communication in terms of linguistic output? For everyday experience, basic level categorization may be sufficient to support our understanding of, and interaction with, the world around us. As the need for different kinds of knowledge changes, the level at which we understand and organize our knowledge may change as well.

2.7 Graded Structure and Prototypes

The graded structure of common taxonomic categories[17] refers, for example, to the representativeness of exemplars within a category (Rosch, Simpson & Miller, 1976). When thinking of a chair, it is likely that most people think of something similar to a desk chair rather than a rocking chair or highchair. In this case, a desk chair may seem more representative or typical of the category than other instances. The findings of graded structure for categories are pervasive within the categorization literature. Murphy (2002) makes this point quite clearly, "This kind of result is extremely robust. In fact, if one compares different category members and does *not* find an effect of typicality, it suggests that there is something wrong with - or at least unusual about – the experiment" (italics in original, p. 24). Before discussing the graded structure of action categories, I will briefly mention three phenomena where the graded structure of categories has been demonstrated: judgments or ratings of the representativeness of category exemplars, judgments of category membership and production of exemplars based on category labels.

2.7.1 Representativeness

Rosch (1975) carried out a series of experiments to investigate the internal structure of categories. In the first experiment, she presented subjects with exemplars from 10 superordinate categories together with the superordinate labels (e.g. furniture, sport, bird, etc.). Subjects were instructed to judge how good an example each exemplar was of the superordinate category indicated by the label. The subject ratings of the exemplars showed that exemplars varied in their representativeness. For example, while a chair was rated as being a very good example of furniture, a shelf was rated as being less representative. If category exemplars vary in their representativeness, it should affect the time it takes to make decisions about category membership.

2.7.2 Judgments of Category Membership

The findings above suggest that the structure of common taxonomic categories is graded with respect to the representativeness of the individual exemplars that are included in the category. A further aspect of graded structure has to do with the extent to which subjects view the category membership of different exemplars. In their classic paper, Rips, Shoben and Smith (1973) let subjects verify the category membership of exemplars from different categories. They found that subjects were quite fast at verifying the category membership of highly typical exemplars (e.g., a robin is bird) and slower at verifying atypical exemplars (e.g., a chicken is a bird).

---

[17] I am restricting the initial discussion of graded structure to common taxonomic categories. Although graded structure has also been demonstrated for ad hoc categories (Barsalou, 1983), the effect of typicality as a determinant of graded structure differs for ad hoc and common taxonomic categories (Barsalou, 1985).

Rips et al. (1973) also found a high correlation between the typicality of category exemplars and the "distance" to the category label. In this case, distance was measured by having subjects rate the pairwise similarity between category exemplars. The more typical an exemplar was of the category, the more similar it was to other highly typical exemplars and thereby "closer" to one another in psychological space. Rosch (1975) also found that the highly representative (typical) exemplars were categorized faster than exemplars with low typicality ratings.

Judgments of category membership have also been used to evaluate graded structure in relation to the "borders" of categories. In this case, subjects were unsure about the category to which an exemplar belonged. For example, Hampton (1979) found disagreement *between* subjects about the category membership of atypical exemplars of the category. The results indicated that there was no clear division between the borders of categories. McCloskey and Glucksberg (1978) also demonstrated an inconsistency *within* subjects when the subjects were asked to perform the same categorization task a few weeks later. For atypical exemplars (olive as a fruit), subjects were more inclined to change their minds from one session to another in contrast to highly typical exemplars. This shows that typicality effects are not simply a result of differences between individuals. These two aspects of typicality (representativeness and decisions about category membership), although related, have been shown to represent two separate psychological processes. See, for example, Murphy and Ross (2005) for a discussion of the role of these two aspects in category-based induction.

2.7.3 Exemplar Production and Typicality

Exemplar production refers to the generation of category exemplars in response to a category label, e.g., *furniture*. Battig and Montague (1969) obtained exemplar production norms for 56 categories by asking subjects to generate as many examples as they could within 30 seconds for each of the 56 category labels (e.g., units of time, four-footed animals, precious metals, birds, clothing, fish, flowers, furniture, sports and vegetables). The resulting data revealed different production frequencies for the listed category exemplars. For example, "robin" was listed by a vast majority of the subjects in relation to the category label 'bird' and pelican was listed by few of the subjects. Another finding was that exemplars with high production frequencies were also among the first items listed. These results preceded the work of Rosch (1975) and Rips et al. (1973) who developed the notion of typicality. In retrospect, however, the production frequencies indicated quite clearly the typicality of exemplars in the various categories.

A more direct assessment of the relation between typicality and exemplar production frequencies was performed by Hampton and Gardiner (1983). If some category exemplars are more typical and thereby potentially more salient in memory, then the more typical exemplars should also be produced more frequently and prior to atypical exemplars in an exemplar listing task. In addition to typicality and production frequency, Hampton and Gardiner (1983) included the variable of familiarity in order to assess the relation between the three variables. Subjects provided ratings of typicality and familiarity for exemplars from 12 categories in the original Battig and Montague study. Familiarity in this study was assessed by asking subjects to indicate how familiar they were with the meaning of each word (exemplar). For the production frequencies, additional subjects were instructed to produce exemplars in relation to the 12 category labels.

Regarding the relation between production frequency and typicality, Hampton and Gardiner (1983) found a significant inverse correlation between typicality and production frequency (-.63) when the effect of familiarity was held constant. The reason the correlation was inverse had to do with the fact that highly typical exemplars were assigned to the low end of the rating scale. The important finding here is that there was a clear tendency for highly typical exemplars to be listed by subjects in the exemplar production task. Atypical exemplars were listed less frequently by the subjects.[18]

2.7.4 Prototypes and Determinants of Graded Structure

The results from the above mentioned studies show that the graded structure of categories can be demonstrated by ratings of representativeness (goodness-of-exemplar), judgments of category membership and exemplar production frequencies. None of the above mentioned studies, however, addresses the issue *of what determines the graded structure* of common taxonomic categories such that some exemplars of a category, in contrast to other exemplars, are deemed more representative of the category, are judged to be clear members of the category and are more frequently listed as exemplars of a category.[19]

---

[18] I should point out that there are apparently different meanings attached to the notion of typicality. Typicality appears to be used as a synonym for graded structure (Barsalou, 1985) but it is also used as a measure of graded structure in e.g., typicality ratings (Hampton & Gardiner, 1983) and exemplar goodness ratings (Barsalou, 1985). The upshot of this is that typicality as a measurement can sometimes be used to demonstrate the phenomenon of typicality. I think this usage is problematic when discussing the factors contributing to the graded structure of categories. I have tried to avoid referring to graded structure as typicality and instead refer to typicality as a measurement or variable. I will attempt to make clear when I use typicality to refer to something other than a measurement or variable. This is an issue that needs clarification in future work.

[19] This issue is similar to the question of category coherence mentioned previously in the chapter. Although category coherence and graded structure are related, the notion of category coherence describes the *general* tendency to classify certain objects as belonging to a specific category whereas the notion of graded structure refers more specifically to the *variation among category exemplars* according to their representativeness (or typicality) in relation to a specific common taxonomic category.

One influential idea about what determines the graded structure of categories was proposed by Rosch and Mervis (1975). According to them, different category exemplars vary in their representativeness (or typicality) on the basis of their *family resemblance* to the other category exemplars, or to a prototypical representation of the category. The resemblance of exemplars to one another or to a category prototype is assessed by the extent to which different exemplars within a category share similar features *and* by the extent to which exemplars share features with category *non*members. In this sense, highly typical members will share many features (properties or attributes) with one another and share few properties with category nonmembers. In contrast, atypical members will share few features with one another and share more features with category *non*members.

The notion of family resemblance is closely tied to the idea of a central instance or best example of a category, i.e., a prototype.[20] Rosch (1978) summarily describes the notion of a prototype in the following way, "In short, prototypes appear to be just those members of a category that most reflect the redundancy structures of the category as a whole" (p. 37). A prototype consists of a summary of (weighted) features that occur among exemplars that are judged as being highly typical of a category. In this sense, a prototype does not have to be a concrete instance, i.e., an instance that we have experienced. A category prototype can also reflect the *central tendency* of a category, where the central tendency represents the average value of category instances (e.g., Hampton, 1979; Barsalou, 1985). The central tendency of a category can then be used to categorize objects by comparing them with the central tendency (or family resemblance) of a category.

Central tendency appears, however, not to be the only determinant of graded structure for common taxonomic categories. Barsalou (1985) found that ideals and frequency of instantiation also play a role.[21] Ideals are features of exemplars that have to do with the specific goal(s) that might be associated with a category. For example, for *vehicle*, subjects were asked to rate how efficient each exemplar was as a means of transportation. Frequency of instantiation, on the other hand, was measured by how frequently subjects thought they encountered an exemplar as a member of the category. One measurement of graded structure in Barsalou's (1985) investigations was goodness of exemplar ratings. In contrast to ideals and frequency of instantiation, Barsalou found that central tendency was more predictive of the graded structure in common taxonomic categories. Ideals and frequency of instantiation, however, were

---

[20] Not to be confused with prototype theory or processing models of categorization. See e.g., Hampton (1995) for a specific model of prototype theory.

[21] A further purpose of Barsalou's study was to investigate potential differences between determinants of graded structure for goal-defined and common taxonomic categories. I have restricted my discussion of Barsalou's results to common taxonomic categories. While acknowledging the role that goals may play in the graded structure of actions, it is not an issue that is addressed in this book.

also significant predictors of graded structure. If central tendency plays a significant role as a determinant of graded structure, a reasonable follow-up question is what factors determine the central tendency of common taxonomic categories.

According to Barsalou (1985), the central tendencies of common taxonomic categories reflect the structure of the environment in the sense proposed by Rosch, Mervis et al. (1976). Objects that are similarly structured according to their shape are perceived as belonging to the same category of objects. The perceptual similarities among objects provide a reliable basis for determining the central tendency of object categories and thereby determine the representativeness of category members and category *non*members.

2.7.5 The Graded Structure of Action Categories

Since natural actions have a spatiotemporal form, and this form can be used to distinguish different actions from one another, it is reasonable to suggest that actions may be deemed more or less typical of category on the basis of the perceptual similarity among exemplars of action categories. The findings from Giese and Lappe (2002) support the idea of action prototypes and accompanying typicality gradients.

Giese and Lappe (2002) (see also Giese, 2002) used a method of spatiotemporal morphing to investigate the effects of varying spatiotemporal patterns on classification, ratings of naturalness and judgment of optimal speed for the actions. They used two sets of displays of biological motion. The first set consisted of four actions depicting different patterns of locomotion (running, walking, limping and marching). These actions tend to be quite similar (SIM) in terms of their spatiotemporal patterns. The second set of four actions depicted actions that had quite different (DIF) spatiotemporal patterns (walking, aerobics, knee bends and boxing). The four actions from each set were used as prototypes from which spatiotemporal morphs were created between the actions. The morphed actions were created by applying a technique called “spatiotemporal morphable models” (STMMs) to generate “new artificial biological movement patterns by *linear combination* of prototypical example movements” (p. 1848, italics in original). The result of the application of the technique was a metric linear space defined by the weights of the linear combination of the trajectories of the prototypes in space-time. Within each set (SIM and DIF), the metric linear space (Figure 2.3) contained 52 actions, including the four prototypes. Given this technique, Giese and Lappe were able to determine the extent to which the spatiotemporal pattern of a specific prototype generalized to, for example, the classification of the neighboring morphs.

When subjects were presented with the two sets of actions, Giese and Lappe (2002) found that the closer a morph was to a nearby prototype the greater its probability was of being classified as that prototype. As the morphs were varied according to the linear weights, subjects seemed to view the gradual change in the perceptual similarity between the morphs and the prototypes. This result was obtained for both the SIM and DIF sets of actions. Despite the fact that subjects could classify the morphs from the DIF set actions as being more similar to a specific prototype, the DIF morphs, in contrast to the SIM set, were perceived as being unnatural.

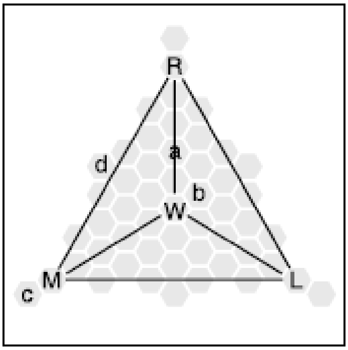


Figure 2.3.[22] A 2D representation of the 4D pattern space of motion morphs from Giese and Lappe (2002). The four prototypes of walking, limping, marching and running are represented by the four letters W, L, M and R respectively. Each hexagon represents an action in the metric space. The letters a, b, c and d represent the combinations of different prototypes and weights. The lines represent morphs between two prototypes.

A further result regarding the classification of the actions concerned generalization fields, which were defined by “the area in the pattern space for which patterns are classified as the same biological motion percept” (Giese & Lappe, 2002; p. 1853). The action of walking appeared to be more similar to the other actions than they were similar to one another as shown by significantly larger generalization fields for “walking” than for the other actions.

Perhaps the most interesting result from Giese and Lappe (2002) for the work presented here has to do with the ratings of naturalness for morphs in relation the prototypes from which they were generated. The issue here was whether the linear combinations of weights of the different prototypes would lead to morphs that are viewed as gradually changing in naturalness or whether subjects would view the spatiotemporal differences between morphs in a more discrete or categorical way, indicating sharp borders between the different action prototypes. This is also another way of viewing the generalization fields of the different prototypes. For the SIM set, the results showed that the perceived naturalness for the morphs was not significantly lower than the perceived naturalness for the prototypes. This indicates that there was a gradual transition in naturalness for the morphs between the prototypes. Of particular

---

[22] Reprinted from *Vision Research*, 42, Giese and Lappe, Measurement of generalization fields for the recognition of biological motion, pp. 1847-1858, 2002, with permission from Elsevier.

interest was the finding of a smooth interpolation between the actions of walking and running and between running and marching. In contrast to these results for the SIM set, the naturalness ratings for the DIF set revealed low naturalness ratings for the morphed patterns in relation to the prototypes. Combining the spatiotemporal patterns of walking, aerobics, knee bends and boxing led to unnatural looking actions.

The spatiotemporal properties of actions give rise to the visual form of natural actions. This visual form can be used to determine the similarity of actions. The extent to which actions have a similar form can indicate categorical divisions between action categories. What are the implications of this for determining a basic level for action categories? One previously mentioned difference between basic and subordinate level categories concerns the greater similarity of subordinate level objects relative to the between-category similarity for basic level objects. If this also holds for categories of natural actions, then the results from Giese and Lappe (2002) suggest that running and walking might be regarded as subordinate level actions, where the category of *locomotion* represents a basic level category. If running and walking represent basic level actions, then we should see greater discontinuity between classification judgments for the morphs between these actions as well as greater differences between ratings of naturalness for the prototypes and the morphs between them. This result was not found. Giese and Lappe were also surprised at this result given the findings of previous research indicating a more discrete phase transition between running and walking (Diedrich & Warren, cited in Giese & Lappe, 2002; see also Hoenkamp, 1978).

Some caution should be observed when drawing conclusions about the basic level of action categories on the basis of the results from Giese and Lappe's (2002) results. First, the actions used in the SIM set consisted of running, walking, marching and limping. As actions of locomotion, the actions are very similar to one another. Marching and limping can reasonably be considered as manners of *walking*. In that sense, it is not surprising that walking had a large field of generalization. Jumping, skipping, crawling and leaping are also arguably examples of locomotion. The question here concerns the extent to which the perceived properties of the morphs would be predicted on basis of the perceived properties of the prototypes for theses instances of *locomotion*. In contrast to the actions in the SIM set, the actions in the DIF set were very different from one another. While boxing can be considered a social action (Dittrich, 1993) and walking is an instance of *locomotion*, it is difficult to find an appropriate category for aerobics and knee bends. One suggestion is that they be viewed as instances of *exercise*. It should be pointed out that Giese and Lappe (2002) did not have the investigation of the structure of hierarchical levels for actions as a specific purpose in their experiment.

The generalization fields indicate graded structure for the category of *locomotion*. Although each action was represented as a prototype in the metric space, there was a much larger generalization field for walking than for the other actions. More of the morphs between the prototypes were classified as instances of walking than as instances of the other actions. In other words, the surrounding morphs were more typical or representative of walking than the other actions. In this sense, walking could be viewed as more prototypical of *locomotion*. Giese and Lappe (2002) suggest the following explanation of their finding; "Walking might be, in the metric defined by the features extracted by the visual system, more similar to most points of the generated pattern space than the other prototypes" (p. 1853). In terms of central tendency, the generalization field for walking indicates that walking has a higher average similarity to the other actions than any other single action and its similarity to the other actions of locomotion. However, a further criterion of central tendency is an action's, or prototype's, average *dis*similarity to contrast categories. For example, if walking has a high average similarity to the other locomotion exemplars, then it should have a high average *dis*similarity to exemplars from the members of contrast categories. This additional aspect of family resemblance was not included in Giese and Lappe's (2002) study.

2.8 Conceptual Spaces and Action Categories

A recent representational format for action categories has been proposed by Gärdenfors (2007).[23] Based on his notion of conceptual spaces (Gärdenfors, 2000), action categories can be represented as convex regions in a conceptual space. According to Gärdenfors (2007), "[a] *convex* region is characterized by the criterion that for every pair of points v_1 and v_2 in the region all points in between v_1 and v_2 are also in the region" (p. 173). The implication of this notion of convexity when applied to action categories is that if two actions are categorized as exemplars of the *same* category, and they occupy separate points in a convex region, then any action exemplar occupying a space between them will be categorized as belonging to the same category. It is important to note that this view takes the context of categorization into account by stipulating that the quality dimensions of actions determine the basis for assigning properties to actions as well as determining the relations among the properties. In this sense, different contexts, perhaps defined by different goals or other situational factors, will lead to the use of different quality dimensions and thereby different regions of convexity. A further central aspect of quality dimensions according to Gärdenfors (2000) is that they should be viewed as *geometrical* structures, and as such we can view objects/actions as being psychologically *closer* (more similar) or *further* from one another (less similar) in a vector space.[24]

---

[23] It should be noted that an additional focus in Gärdenfors (2007) is on the representation of the functional properties of objects. This is not an issue specifically addressed in this book.

[24] The notion of convexity in Gärdenfors (2000; 2007) is also apparent but not explicitly mentioned in the computational model in Giese (2002). I should also state that the ideas of Gärdenfors and Giese were developed independently of one another.

2.8.1 The Role of Forces

Regarding the quality dimensions of actions, Gärdenfors (2007) proposes that action representations are fundamentally determined by the *forces* that generate them. The basic idea here is that different force patterns are involved in the production of different actions. Previous research on the relationship between the dynamics of human movement and the kinematic patterns that arise as a result of dynamic constraints of the human body show that human observers are sensitive to the underlying forces involved in human movement and even our interaction with objects (Runeson & Frykholm, 1983; Pollick & Kourtzi, 1998). Further support for the role of forces in action perception comes from findings on representational momentum when people view static images of implied human movement (e.g., Freyd, 1983, 1987; Shiffrar & Freyd, 1993; Kourtzi & Shiffrar, 1999; Kourtzi & Kanwisher, 2000; Verfaillie & Daems; 2002). For example, Verfaillie and Daems (2002) found significant long-term priming effects when subjects were primed with a short animated sequence of human movement and then tested with static images of possible future postures which were consistent with the previously viewed movement sequence. This indicates that memory effects of human movement include the dynamic properties associated with motion constraints involved in the future positions of the human body in time.

In terms of the basic components mentioned by Tranel et al. (2003), the *manners* in which different actions are performed reflect the differences between the spatiotemporal trajectories of body parts. Variations in spatial trajectories can be used to make coarse distinctions between actions as well as making more fine grained distinctions. For example, Klatzky, Pellegrino, McCloskey and Lederman (1993) presented subjects with verb phrases for arm- and hand-related actions that involve objects (brush hair, chop onions, catch a ball, hammer a nail, etc.). The subjects were then given the task of rating the actions according to *effector size* (Which limbs are involved in the action?), *amount of limb* (How much of the limb(s) is(are) in motion during the action?), *amount of surface contact area* (How much of the limb comes in contact with some object?), *distance moved* (How much does the limb move through space as the action is performed?), *resemblance to grasp* (If the hand was used, does it grasp the object or operate without grasping?) and *amount of force* (How much force is used in the action?). Correlations between the ratings and a subsequent factor analysis showed that *limb*, *distance* and *force* were highly correlated and that "nearly 80% of the variance was accounted for by factors that seem to represent arm movement/force and hand configuration" (p. 297). This shows that spatial trajectory as indicated by movement and force is a part of the knowledge subjects have about arm-related interactions with objects. In a second study, Klatzky et al. (1993) also showed that this knowledge can be used by subjects to create categories of the different arm-related actions. The categories formed by the subjects reflected the identified factors from the first study, indicating a clear division according to involvement of the arm or hand and the extent to which the actions were related on the dimensions of force, limb and distance. Perceived forces in arm- and hand-related actions contributed significantly to category distinctions as well as the creation of action categories.

While forces constrain action production and recognition and may even be perceptible properties of actions, this does not rule out previously mention influences of body parts and the spatiotemporal patterns that arise from their movement. As Runeson and Frykholm (1983) suggest in their principle of the "Kinematic Specification of Dynamics" (KSD), the kinematic patterns of human movement reflect the dynamics that constrain those patterns. The upshot then is that both spatial and temporal aspects of human actions constitute quality dimensions within an action space. (See, e.g., Thornton (2006) and Giese (2006) for brief reviews of temporal and spatial manipulations of action sequences using point-light displays of biological motion.)

2.8.2 Action Spaces, Prototypes and Graded Structure

If one adopts the notion of conceptual spaces as convex regions in a vector space, then the previously mentioned findings of graded structure and prototypes for categories are to be expected (Gärdenfors, 2007). Regarding the domain of colors, Gärdenfors asserts, "For example, if colour concepts are identified with convex subsets of the colour space, the central points of these regions would be the most prototypical examples of the colour" (p. 176). Although we know relatively less about the quality dimensions that characterize action spaces and the ways in which the quality dimensions can be combined to create conceptual spaces for actions, the previously mentioned results from Giese and Lappe (2002) strongly suggest that a metric representation of action categories captures important psychological findings. For example, the naturalness ratings of the *action morphs* provided by subjects could be reliably predicted on the basis of the naturalness ratings of each *action prototype* and the respective weight of the prototype in the linear combination (Giese & Lappe, 2002). This shows that the convexity of the action space represented by action morphs that lie between the action prototypes is reflected in the naturalness ratings provided by the subjects.

Further support for the representation of actions according to a psychological space comes from Pollick, Fidopiatis and Braden (2001) where they recorded the movements of different kinds of tennis serves (flat, topspin and slice). The vectors representing the different motions of body parts, as well as the tennis racket and ball, were calculated on the basis of points attached to the body of the person performing the different serves. The average movement associated with each kind of serve as well as the grand average for all three kinds of serves were also derived from the movement data. In this case, the grand average of the three kinds of serves represented the prototype for a tennis serve. By using the difference between the movement for a specific kind of serve and the grand average, Pollick et al. were able to produce spatial exaggerations for each kind of tennis serve. The main questions in their study were whether increasing exaggerations would show a corresponding accuracy in categorization judgments and whether the varying exaggerations would lead to differences in dissimilarity judgments such that the differences would be reflected in the structure of an obtained 3-D psychological space. With the exception of the slice serve, the results showed that categorization judgments improved as the exaggerations moved further away from the grand average. For the dissimilarity judgments, the results showed that increasing the spatial exaggerations in the movements resulted in a corresponding difference in the distance between the movements in psychological space.

While the results from Pollick (2004) and Pollick et al. (2001) showed that a radial structure of the psychological space for the different tennis serves could be obtained, the results also showed that the style exaggerations lead to better categorization performance than the style prototypes. These results indicate that categorization performance is not always facilitated by proximity to a category prototype. It appears that spatially exaggerating the specific movements of an action in a direction *away* from the central tendencies of other similar actions can increase the spatiotemporal distinctiveness of that action and thereby make it easier to distinguish it from the other similar actions, i.e, tennis serves. In line with the findings from Murphy and Brownell (1985), small differences between exemplars may be the basis on which subordinate exemplars are distinguished from one another. When these differences are exaggerated, they lead to greater distinctiveness for subordinate level exemplars. In this case, exaggerated exemplars of subordinate level categories will be more distinctive in relation to other subordinate level exemplars that share features with members of contrast categories. Recognition of subordinate level exemplars is made more difficult due to the fact that they may share certain properties with members of contrast categories. For example, different kinds of *cars* may be more confusable than different kinds of *vehicles*. See Goldstone (1996) for more findings regarding this issue.

The results from Pollick et al. (2001) should not be understood to mean that spatiotemporal exaggerations will benefit the categorization of actions for all such categories. The Pollick et al. results are somewhat limited in terms of their generalizability to other potential levels in a category hierarchy for actions. The different tennis serves were very similar and the subjects who viewed the displays were classified as intermediate level players. The effect of spatiotemporal exaggerations may vary as a function of skill level among players or even among expert observers of the game. A player with a very high skill level may not benefit from the exaggerations to the same extent as a novice or an intermediate level player. As previously mentioned, level of expertise affects the ability to categorize objects/actions on a finer grained level (Tanaka & Taylor, 1991).

On the basis of the above mentioned findings, it appears that a geometric representation of actions defined by the derived distances between actions is reflected in behavioral measures of categorization. Given the salient spatiotemporal aspects of dynamic human movement, it seems reasonable that a geometric representation of the domain of human actions could be extended to action categories beyond the ones investigated in the previous studies. If the psychological organization of other action categories can be captured by the psychological distance in a geometric space, then perhaps this will be reflected in the mental lexicon as well. This issue will be explored in the next chapter.

2.9 Summary

The visual shape of actions appears to be a highly predictive feature of action categories. Part of this predictive quality of the visual appearance of actions appears to be tied to the production of human movement. This indicates one important difference between human action perception and the perception of the motion of objects. Human movements signal social interaction whereas object motion may or may not signal functional interaction. In the case of artifacts, the movement of parts of objects signals functional interaction.

In regards to the issue of whether action concepts constitute a different “kind” of concept (Medin et al., 2000) in contrast to object concepts, results from Shiffrar, Lichtey and Heptulla-Chatterjee (1997) suggest that action perception and object motion perception may depend on different motion integration mechanisms. This processing difference is one criterion mentioned by Medin et al. for distinguishing between kinds of concepts.

It is surprising that there is little mention of action categories among the myriad of research on concepts and categorization. Review articles (e.g., Medin et al., 2000) and books on concepts and categorization are remarkably silent about action concepts. The obvious question is: why is this so? One obvious reason is that there is little research done on action concepts. The follow-up question is then: why is there so little research on action concepts? Actions are difficult to study. They are dynamic and easily confounded with other variables. A related reason has to do with the fact that actions are relational in nature. They include information about relations between objects, i.e., a human and another human, or a human and an object. The major point here is that we know relatively little about action concepts and categories in themselves, and we know even less about the relation between action concepts and how they may differ from, or be similar to, other kinds of concepts (Medin et al., 2000).

To what extent is action recognition constrained by the psychological organization of action categories according to hierarchical levels, basic level effects and graded structure? Action recognition will likely be influenced by other structural and functional factors, but the purpose here is to first gain some understanding about action categories in relation to classical findings in the categorization literature. I realize, however, that the "classical findings" have been subjected to important qualifications regarding the multiple functions of concepts (Solomon, Medin & Lynch, 1999).

Several findings (Dittrich, 1999) point to the role of higher level categorical knowledge in the perception and recognition of actions, for example, findings indicating that action categories have graded structure (Giese & Lappe, 2002), prototypes (Giese & Lappe, 2002; Pollick et al., 2001) and hierarchical levels of organization (Dittrich, 1993). There is no specific evidence, however, for basic level effects in action categories, although there is some evidence that can be interpreted as an indication of such effects.

A further aspect of the higher level knowledge associated with action perception is the idea that categorical knowledge may be reflected in a multi-dimensional space where the psychological distance between actions can be used to see clusters of action categories, which may indicate categorical breaks between different kinds of actions. Given the extent that the spatiotemporal dimensions of actions indicate perceptual salient qualities of actions and their dynamic constraints, it may be the case that the psychological organization of action categories is stable across different languages and cultures.

Chapter 3 - Hierarchical Structure of Action Categories[25]

In addition to recognizing objects, a significant aspect of our daily activity involves perceiving and recognizing the actions of other individuals. Not only do we see certain things as CUPS, BOOKS, DOGS, CARS, APPLES, etc., but we also see various patterns of movement as RUNNING, WALKING, JUMPING, THROWING, etc. Furthermore, the ability to recognize actions and events would seem to be a basic cognitive function given the fact that we live in an environment that is largely dynamic with respect to our own movements within it, including interactions with objects and people, and with respect to our perception of the movements of others and their interactions with other people and objects.

Much of the research dealing with the connection between what we see and the subsequent lexicalization of our percepts into concept hierarchies has been mainly addressed from the perspective of the object-noun relationship. The primary categorical domains that have been investigated have been those dealing with natural kinds and artifacts. (See Medin and Smith (1984) and Mervis and Rosch (1981) and Komatsu (1992) for reviews of relevant theories.) Dittrich (1993) presented results concerning the categorization of actions based on biological motion sequences.

A widely held assumption in accounts of categorization is the relation between exemplars and their superordinates and that people have access to this relation in the context of categorization studies. If presented with a superordinate concept, subjects have no difficulty producing exemplars in relation to it. This type of task can be used to investigate the relation between different levels of a concept hierarchy, e.g., superordinate => basic level, and basic level => subordinate. In this sense, one gains insight concerning the kinds of exemplars that are produced in relation to a given superordinate. The work reported here is intended to extend the research on action

---

[25] This chapter has been previous published as: Hemeren, P. E. (1996). Frequency, ordinal position and semantic distance as measures of cross-cultural stability and hierarchies for action verbs. *Acta Psychologica*, 91, 39-66.

categories from the perspective of the lexical items generated by subjects when given a description of a general (superordinate) concept based on perceptual criteria for the basic level.

Given the previous work on object categorization and the cognitive primacy of basic level categories and the significant role perception plays in the formation of such categories (e.g., Rosch et al., 1976), it may very well be the case that basic level perceptual criteria can be applied to, at least, a general domain of action categories. If there is some middle level for action categories that is similar to the middle level for object categories in terms of what gives rise to them, then one would expect a similar basis for perception in determining this middle level in the concept hierarchy. The strategy here is to give subjects the general perceptual criteria and have them list the actions that meet the criteria. These actions then are interpreted as being good candidates for the basic level.

I must emphasize that in the sense that the task used here is a categorization task, it is quite different from giving subjects certain action sequences and asking them to identify the sequences (identification task) or asking them to verify that an action belongs to a previously presented word that denotes a given concept (verification task). The difference is roughly this: in the studies reported here, the direction of the task is from general perceptual criteria to specific lexical concepts, whereas the perhaps more typical method used in categorization studies involves going from specific perceptual input in the form of an image of some kind to lexical concepts that name the presented objects or actions. Both methods have their virtues in the attempt to empirically establish a hierarchical structure for action categories. The hope is that the two methods will independently converge on a similar lexical structure for action naming.

This chapter presents normative data concerning the response frequencies[26] for a general class of actions. In an action listing task, response frequencies were generated by a native English speaking group and a native Swedish speaking group. The results indicate that the general perceptual criteria for the basic level can be applied to action categories and the varying response frequencies demonstrate graded structure within the general class of actions. While the most frequent verbs are those that might best be

---

[26] I will use the term response frequency throughout the remainder of the paper to refer to the total number of times a given item occurs across subjects on a free listing task as used by Battig and Montague (1969). Other people have referred to this measure differently; associative frequency (Hampton & Gardiner, 1983); item dominance (Mervis et al., 1976); production frequency (Malt & Smith, 1982) and output dominance (Barsalou, 1985). There are two reasons for my choice of the term response frequency over the others. (1) It is the term that seems to be closest to the original Battig and Montague usage, and (2) it seems to be the most parsimonious description of the dependent measure, i.e., the individual subjects were simply providing responses with respect to a given task. Furthermore, the subsequent tallies of the responses were frequencies and not tallies of dominances per se, although dominance could be used as a description of the varying degrees of response frequency.

considered as basic level, the subordinates of these basic level verbs occur much less frequently and much later on in the lists. In order to establish a further measure of basic levelness, the distribution of the response frequencies between the two language groups were compared in order to determine the cross-cultural/linguistic stability of the most frequently listed actions. The amount of agreement between the two groups suggests a high degree of stability across the languages for the most frequently listed actions. In addition to the response frequencies, multidimensional scaling solutions based on the ordinal structure of the lists were performed in order to answer questions concerning the semantic groupings of the words in the lists and the cross-cultural stability of these groupings as well as the overall stability of the response frequencies.

Concerning the usage of the terms “concept” and “category,” I will be loosely following Medin’s (1989) distinction that a concept is an idea that includes all that is characteristically associated with it and that a category is a partitioning or class to which some assertion or set of assertions might apply (cf. chapter 2). I will refer to concepts as that which becomes lexicalized in the form of nouns and verbs.

3.1 The Basic Level: Cognitive Primacy and Perception

A salient finding in categorization research is the cognitive primacy of the basic level as compared to the superordinate and subordinate levels of categorization (Rosch, Mervis et al., 1976; Rosch, 1978; Murphy & Smith, 1982). The basic level is the primary level at which category differentiation reflects the natural divisions of attribute clusters found in the environment.

One important constraining factor in the acquisition and formation of categories is perception, and on the basic level this is particularly so. Rosch, Mervis et al. (1976), Rosch (1978), Neisser (1987), Mervis and Crisafi (1982), and Tversky and Hemenway (1984) express a general consensus that there are two unique properties of the basic level: (1) Members of basic level categories are similar in overall shape and (2) similar with respect to our interactions with them, i.e., they have similar functions as in the case of artifacts. Mervis (1987) refers to these two properties as constituting the ”shape/function principle.” Accordingly, much of categorization, but by no means all, is a result of the application of this principle. The shape/function principle is largely perceptually driven in the sense that the visual shape of an object can be obtained by looking at it. Function, on the other hand, may not be so readily analyzable with regard to perception. However, although one may not be able to tell what the function of an object is by looking at it, some insight concerning function can be gained by interacting with the object or by watching someone else interact with it. From the perspective of these results, it would seem that perception ought to be an appropriate starting point from which to investigate whether or not action verbs can be generated on the basis of perceptual criteria.

3.2 Action Categories and Biological Motion

Using perceptual criteria to get subjects to generate lists of actions will only be effective if there is a class of actions to which such perceptual criteria apply. Actions of bodily movement seem to be such a class. As Miller and Johnson-Laird (1976) put it:

> Not only are verbs of motion ontogenetically primary, but their meanings have a strong perceptual basis – a correlation that can hardly be coincidental. When someone cogitates or acquiesces or experiences it is not clear just what perceptible signals of those "activities" he will transmit, but when he runs or jumps or climbs there is little question. (p. 527)

Furthermore, there is reason to believe that such actions are perceptually basic in that they can be recognized quickly, though not so much on the basis of context as on the pattern of movement of the parts of the body. A prime example of this perceptual basis can be found in the work of Johansson (1973; 1975) and his colleagues (Kozlowski & Cutting, 1977; Cutting, 1981; Runeson & Frykholm, 1983). Johansson (1973) describes a study in which he placed small lights on the joints of a person who performed various actions. The subjects in this study were readily able to discern a number of biological motions, e.g., running, cycling, climbing, and dancing, by simply viewing the resulting flow patterns of the lights. The demonstration of this patch-light technique has two interesting ramifications for the categorization of actions. In one sense, the patch–light figures contain very little contextual information. But in another sense, they contain a great deal of kinematic and dynamic information in the flow patterns of the lights. Secondly, subjects were very good at recognizing a given action on the basis of only viewing a few frames from the motion sequences. The results of Johansson and his colleagues suggest that perceptual criteria may also provide a basis for action categories. (The patch-light technique will be described in further detail in chapter 4.)

Regarding the issue of context sensitivity mentioned above, it should be added for the sake of clarity that context sensitivity has been demonstrated for action categories (Vallacher & Wegner, 1987). It may even be the case that such categories are even more context sensitive than object categories. The suggestion here is simply that there may be a group of actions that is much less context sensitive than other kinds of actions and that this may depend on the extent to which social setting and perception mutually constrain the categorization of actions. Although it is not the case that what distinguishes a certain group of actions from other actions in a concept hierarchy is only the degree of perceptual salience, perception seems to be one unequivocal factor in the formation of action categories at a middle level in an action concept hierarchy. It seems quite likely that the function of certain actions, i.e., the fulfilling of some goal, in a social setting is also important. This is in accordance with the similarity/function principle mentioned above. Evidence of the convergence of operations on some middle level of the concept hierarchy is also needed in order to principally establish a basic level for actions. The method and analyses described here present a step in this direction.

The idea that there are basic level action categories is not novel. In *Women, Fire and Dangerous Things* Lakoff (1987) asserts, "We have basic level concepts for actions and properties as well. Actions like *running, walking, eating, drinking,* etc., are basic level, whereas *moving* and *ingesting* are superordinate, while kinds of walking and drinking, say, *ambling* and *slurping,* are subordinate" (pp. 270-271). The studies below are intended to investigate these claims by examining the responses of subjects in relation to a free listing task for actions.

3.3 Study 1: Response Frequencies for Action Categories (American Sample)

In this experiment, a free listing task very similar to the one used by Battig and Montague (1969) is used. In their article, however, they included 56 different categories, whereas the present experiment uses only one very general superordinate category. A number of predictions can be made on the basis of the findings and reasoning presented above. First, if subjects are given perceptual criteria for action categories and asked to generate lists according to the general perceptual criteria, subjects should be able to interpret the task as meaningful in the sense that the perceptual criteria apply to actions in a way similar to the object categories from which they were taken. Secondly, certain types of verbs or verb phrases should occur more often than others (graded structure) rather than being evenly distributed among the lists. A further prediction is that the verbs will be similar to the ones mentioned by Lakoff (1987) and Miller and Johnson-Laird (1976). It certainly seems plausible that subject lists would contain varying response frequencies of action words and that high frequencies would be obtained for verbs that denote actions like, *eating, walking, running, jumping,* etc. more so than other more context dependent actions like, *buying a car, teaching, going to a restaurant,* etc. Finally, if high frequency is taken as an indication of basic levelness, subordinates should occur at much lower frequencies. And with regard to their ordinal positions, verbs with high frequencies, assuming varied distribution, should also be the ones that occur earlier on in the word lists. That is, the cognitive primacy of the basic level should also be revealed in terms of the ordinal positions of the verbs in relation to subordinate level verbs.

3.3.1 Methods

3.3.1.1 Subjects

A total of 119 American English native speaking Hope College undergraduates from five psychology classes volunteered 10 minutes of their time to participate as subjects.

3.3.1.2 Materials

The subjects were given a sheet of paper with instructions written at the top. Below the instructions, and on the reverse side of the sheet, were numbered blanks for the subjects to fill in during the timed writing session.

3.3.1.3 Instructions

Writing the instructions for the generation task posed a problem. On the one hand, the instructions had to be easy to understand. For example, I did not want to have to go into an explanation about what the basic level is and how there might be basic level actions. On the other hand, the instructions had to be meaningful and somehow constrain list generation to the realm of actions that met certain perceptual criteria. The perceptual criteria used in the instructions were adopted from Mervis and Rosch (1981) who point out three special properties of the basic level for objects. The first property is that a person uses similar motor actions for interacting with category members. The second property is the similar overall shapes shared by category members, and the third property is a mental image which can reflect the entire category. Since the first property is confined to actions in the service of object function, the criterion "ease of recognition" was used instead in order to maintain the generality of the perceptual criteria. The property that a mental image can reflect the entire category may be viewed as a result of the similar overall shapes of objects. The second and third properties were therefore combined into a single mental imagery criterion. The resulting instructions presented to the subjects were as follows:

> The purpose of this session is to collect verbs that name various actions. You are simply to write down, on the numbered blanks below, words or phrases that names various actions. It is important though that the words or phrases name actions that involve some kind of *bodily activity that can easily be recognized when seen and can be visualized as a mental image.*
>
> You will be given five minutes to write down as many words or phrases as possible that name different actions of bodily activity. Please write neatly. Thank you for your participation. If you have any questions, I will answer them now, but do not mention any possible examples of actions. You can begin when I say "Please begin."

3.3.1.4 Procedure

After all the subjects received a copy of the instructions and numbered blanks, an experimenter read the instructions out loud. No subjects in any of the five classes had any questions.

3.3.2 Results and Discussion

First, words were scored as the same if they were orthographically identical or only varied according to tense. Subjects appeared to have little difficulty in understanding the nature of the task. The mean number of words or phrases per list was 36.36, SD =

10.91, median = 35. The minimum and maximum lengths of the lists were 14 and 72. A total of 920 different words were produced.

In the following analyses of the subject lists, two dependent measures are used. The total frequency (TF) for each word indicates the total number of times a word appeared across the 119 different lists. The second measure is the mean ordinal position (MOP) and represents the averaged ordinal position of a word across all the lists on which the word appeared. See Appendix A for the list of words that have a T F of 3 or more.

The TFs presented in Appendix A confirm the general hypothesis of an uneven distribution of response frequencies for action words (graded structure), i.e., some words are more salient examples of bodily activity than other words. This finding is somewhat trivial when understood in the context of Barsalou's (1987) statement that "every category observed so far has been found to have graded structure." Therefore, the graded structure found here is no exception to the general finding. More interesting, however, is the finding that the words that received the highest frequencies tend to belong to the class of action words mentioned by Lakoff (1987) and Miller and Johanson-Laird (1976). That is, words like RUNNING, JUMPING, SWIMMING, WALKING and EATING occurred more often than more context dependent words like TEACHING, BAKING, WRITING A LETTER, EATING BREAKFAST and ARGUING. The superordinate category, as defined by the perceptual criteria, corresponds well to the kinds of verbs that denote the perceptually salient actions used in studies of biological motion.

In reporting the coefficients for some of the correlation analyses that follow, I will use both the Pearson r and Spearman r_S. The reason for reporting both is due to different views concerning the kinds of analyses that can be used in regard to different scales of measurement.[27] Whereas the measure TF is based on a nominal scale, the MOP is base on an ordinal scale, and some differences may arise as a result of the kinds of analyses that can be performed given the nature of the scales (Stevens, 1951). Another reason for reporting both is that the results presented below are discussed in the context of previous work where the Pearson r was used exclusively. All probabilities associated with the correlations are two-tailed. The assessments of levels of significance for the Spearman r_S are based on Glasser and Winter's recommended t test as cited in Nijsse (1988).

The correlation for the relation between frequency and ordinal position reveals a strong trend for the most frequent items to also appear early on in the subjects' lists. The words included in this correlation had a TF of 20 or more. As the TF increases, the MOP tends to decrease, Pearson $r = -.70$, $p \leq .0001$ ($r_S = -.54$, $p \leq .001$). It appears that there are a few dominant items in memory that get written down first, and

---

[27] See Cliff (1993) for a discussion.

they seem to be the best examples of the general superordinate category as indicated by the significant trend.

3.3.2.1 The Relation Between the Basic and Subordinate Level Items

A qualitative perusal of the Battig and Montague data not only reveals a relation between a given superordinate and the basic level items subsumed by it but it also reveals a relation between some of the basic level items and their subordinates. While there is a striking tendency for basic level items to have a relatively high frequency (and a relatively low rank position where it is shown in their data), subordinate level items had a lower frequency (and a higher rank order). As an example, for the superordinate category A KITCHEN UTENSIL, the most frequently mentioned item was KNIFE. The subordinate items, or kinds of knives, occurring in the lists included, PAIRING KNIFE, CARVING KNIFE, BUTCHER KNIFE, BREAD KNIFE, BUTTER KNIFE, and CAKE KNIFE. Another example is that of AN ARTICLE OF FURNITURE. While the basic level item CHAIR was the most frequent, subordinates occurred less often and had a higher rank position on the lists. The subordinates included were LOUNGE CHAIR, EASY CHAIR, ROCKING CHAIR, ARMCHAIR, HIGH CHAIR, DESK CHAIR, DINNING-ROOM CHAIR, DORM CHAIR, LAWN CHAIR, and RECLINING CHAIR.

Results very similar to those found in the Battig and Montague data appear in the verb data. Presented below are 4 frequently listed verbs and their subordinates. The frequencies (TF) are reported first followed by the mean ordinal positions (MOP). A confirmation of the hierarchical relation between the verbs was obtained by checking the relations via WordNet™ (version 1.4), a lexical database that shows encodings of the hierarchical relations between synonym groups. This was obtained by having the program list the particular ways of RUNNING, for example. And although there are different senses of RUNNING, I chose the sense that seemed to best match the nature of the task given to the subjects. The sense and its definition according to the WordNet™ database are also given.

Running (115: 4.37) (Sense 19, move by running) => jogging (39: 11.87), sprinting (9: 13.89), trotting (5: 16.20)

Walking (99: 7.32) (Sense 3, walk, go on foot, foot, leg it, hoof, hoof it) => hiking (9: 19.11), sauntering (6: 19), strolling (5: 17), pacing (4: 25), hobbling (3: 18.67), limping (3: 32), marching (2: 24), ambling (1: 28), tiptoeing (1: 31), moon walking (1: 22), moseying (1: 6), staggering (1: 13), strutting (1: 26), swaggering (1: 6), stumbling (6: 17.83)

Jumping (92: 7.61) (Sense 4, jump, leap, bound, spring) => hopping (32: 8.47), leaping (18: 18.67), skipping (61: 7.54), bounding (2: 17), hopping on one foot (1: 9)

Talking (56: 13.71) (Sense 4, talk, speak, utter, mouth, verbalize, verbify; (”She talks a lot about her childhood.”) => yelling (38: 17.50), screaming (24: 21.21), whispering

(17: 24.94), speaking (14: 19.50), shouting (11: 25), arguing (2: 12), saying (2: 33.5), preaching (2: 31), articulating (1: 4), discussing (1: 52), mumbling (1: 21), remarking (1: 30), responding (1: 13), stuttering (1: 34), telling (1: 25)

It does not appear to be the case, however, that all such basic level items that have a high frequency also subsume a group of subordinates that have a lower frequency. It may be that some basic level categories have relatively few subordinates. Or it may be the case that access to the subordinates is more constrained in terms of the context that would give rise to them. For example, different ways of walking and talking may be much more prominent in terms of our need to lexicalize them whereas different ways of running and jumping are less prominent.

3.3.2.2 The Instability of Graded Structure

Another aspect of the data that deserves comment is that there are only 5 verbs mentioned by more than half of the subjects (RUNNING, WALKING, JUMPING, SWIMMING and SKIPPING). This relative lack of uniformity across subjects indicates a wide disparity between individuals concerning the relation between the superordinate and its exemplars. Apart from the 5 verbs, subjects do not seem to access similar semantic or categorical domains in relation to the general perceptual criteria. Another way of putting it is that the general structure of the category of actions that are viewed as being subsumed by the perceptual criteria is unstable from subject to subject. This instability, however, is strikingly similar to the lack of uniformity in the categories in the Battig and Montague norms where the average number of items that are mentioned by more than half of the 442 subjects for each of the 56 categories is 3.95. (See also the results for production data in Barsalou (1987)) An additional measure of graded structure is typicality. And although no typicality ratings were gathered here, one could expect that the lack of agreement between subjects would correspond to the correlations mentioned by Barsalou (1987) where intersubject agreement in typicality ratings gathered in numerous experiments hovers around .50.

A few things must be kept in mind when discussing the instability of graded structure as indicated by frequency data. In one sense graded structure is by definition an indication of the instability of category structure between individuals. That is, varying frequencies for the different items indicate graded structure and instability. Some words are listed more often than others. In another sense, however, the relation between graded structure and instability can be viewed as separate notions. For example, if there were no overlap between the words on any of the subject lists, where all words had a frequency of 1, then this would indicate complete instability and no graded structure what-so-ever. Regarding the other extreme where all words were listed by all subjects, one would then have a situation of complete stability and no graded structure. To the extent that graded structure is a function of stability regarding frequency data and assuming a quantitative measurement of graded structure, one can say that as stability increases from 0, there is also an increase in graded structure to a point where graded structure begins to decrease given a continued increase in stability to the extreme where there is complete stability and no graded structure. This, however, is not the case with typicality as a measure of graded structure because every subject contributes a rating to every stimulus item. In this case graded structure is a function of the mean typicality ratings for an item, and instability is viewed as a lack of agreement between subjects for a given item. With typicality ratings, one can still have graded structure even if all subjects agree on all typicality ratings, but then there would be no instability. There will also be graded structure even if subjects are in wide disagreement about their respective typicality ratings. In the first case, frequency, the two notions are dependent on one another, whereas in the second case, typicality, they are treated independently.

Although the frequency data presented here can be viewed as indicating instability as well as graded structure, one must recognize that stability is also present. Briefly, Barsalou (1987) proposed an explanation of graded structure and stability effects that rests on a distinction between context-independent and context-dependent information and various determinants of graded structure, e.g., goals, central tendency and frequency of instantiation. Associated with concepts are two different kinds of information. Whereas context-independent information is necessarily linked to a concept and is activated regardless of different contexts, context-dependent information is only activated given an appropriate context. According to his theory, the extent to which people share context-independent information should give rise to similar concepts in a superordinate->basic level listing task like the one used above. The instability, on the other hand, can be accounted for by the fact that individuals' concepts can vary according to the context-dependent information that may be accessed given their understanding of the task *and* by the fact that not all individuals share the exact same kind of context-independent information.

3.4 Study 2: Response Frequencies for Action Categories (Swedish Sample)

The Battig and Montague (1969) frequencies were collected at the Universities of Maryland and Illinois. There were 270 subjects from Maryland and 172 from Illinois who were given 56 category labels and asked to write down as many items as they could within 30 seconds for a given category. Battig and Montague computed correlation coefficients in order to determine the, in this case, "geographical stability of the response frequencies for the Maryland and Illinois samples." The results revealed strong evidence for geographical stability. Forty-nine of the 56 categories had a correlation coefficient greater than .90.

Using British subjects, Hampton and Gardiner (1983) collected normative data for 12 of the categories used by Battig and Montague. One purpose of the study was to see if there was any cross-cultural variation between the two populations. The resulting comparisons between the response frequencies for the 12 categories revealed coefficients that ranged from a low of .48 for FISH to a high of .91 for WEAPONS. The mean coefficient for the 12 categories was .76, indicating that the categories are rather stable across the two groups, but yet different enough to warrant the collection of separate norms for use with British subjects.

The minimization of cross-cultural differences in categorization is an additional aspect of basic level categories (Medin & Barsalou, 1987; Rosch, 1973). The extent to which the basic level is grounded in perception and by constraints that span the boundaries of cultural differences and context sensitive variables ought to be revealed by the stability of the categories across culture and language. The second study of this chapter was conducted in order to investigate the stability of the action categories across language (and culture). The stability referred to here is that which needs to be accounted for due to the fact that there exists some agreement between the subjects, otherwise there would have been no graded structure. The hypothesis is that similar categories and response frequencies should appear for the same task used in the previous experiment with subjects from a different country who speak a different language. If the verbs with the highest frequencies generated in the first experiment have the quality of being primarily perceptually based in the sense that the pattern of bodily movement is sufficient for recognition and categorization and that actions categorized on the basis of this information are common actions that humans perform, then one would expect a high degree of cross-cultural stability. A group of Swedish students was given a translation of the instructions used for the American group. The results from this group were compared to a sample taken from the American group. The two groups were compared to see if similar action words are also the most frequent for the Swedish group and to see if their ordinal positions were similar as well.

Admittedly, the best test of cross-cultural stability would be to compare two groups that are more different than ones used here. The “best test,” however, is not always the most realistic. The main reason for choosing Swedish as the comparison language is that, next to English, it is the only language that I speak fluently enough to do the kind of semantic comparisons presented here. And even though the two cultures are quite similar, the comparisons should be seen in the context of the British English and American English comparisons in the Hampton experiments mentioned above. The two kinds of English are obviously closer to one another than American English and Swedish. If the results of the comparisons in this experiment are similar to the Hampton results, then there is reason to believe that the notion of stability is just that much stronger.

3.4.1 Method

3.4.1.1 Subjects

Thirty-nine Swedish speaking undergraduate students from an introductory psychology course volunteered 10 minutes of their time to participate as subjects.

3.4.1.2 Materials

The materials were the same as described in Experiment 1A.

3.4.1.3 Instructions

The English instructions were translated into Swedish.

3.4.1.4 Procedure

The procedure was the same as in Experiment 1A. One student, however, had a question concerning the nature of the actions referred to in the instructions. The experimenter re-read the portion of the instructions describing the general class of actions that were to be listed and instructed the subject to write down the actions that best seemed to fit that general description. The subjects were tested in one group.

3.4.2 Results and Discussion

For the English speaking sample, 39 subjects from the group of 119 were randomly selected to be used in the Swedish/English comparisons. The mean number of words per list for the English group was 37.13, SD = 10.58. The minimum and maximum list lengths were 25 and 72 respectively. As in the first study, the TF and MOP for all words were calculated across all the lists. These measures were then compared with the same measures from the larger sample in order to determine the representativeness of the smaller sample. This was done for words with a frequency of 20 or more in the larger sample. The coefficient for the correlation between the TFs was .96, $p \leq .0001$ ($r_S = .89, p \leq .001$), which shows that the smaller sample is representative of the larger sample with regard to the distribution of frequencies. The correlation between the MOPs for the two samples was somewhat lower, $r = .85$, $p \leq .0001$ ($r_S = .79$, $p \leq .001$). This indicates that word position is less stable than the distribution of response frequencies across the two groups. As in the analysis for the large sample above, a coefficient was calculated for the correlation between TF and MOP in the smaller sample. The coefficient for this correlation for words that had a TF of 10 or more was -.63, $p \leq .0001$ ($r_S = -.46$, $p \leq .01$). As in the large sample, there is a significant trend of decreasing ordinal position as the TF increases.

3.4.2.1 The Relation Between the Basic and Subordinate Level Items

For the Swedish speaking sample, the mean number of words per list was 41.56, SD = 12.62. The minimum and maximum list lengths were 20 and 67 respectively. A comparison of the list length means for the Swedish and English samples showed no significant difference, F(1,76) = 2.83, $p > .05$. Subjects in both groups generated the same average number of words per list. As in the two English samples, there is a trend in the Swedish results for words with the highest frequency to also be listed earliest in the lists. The coefficient for this correlation (for words with a frequency of 9 or more) was -.69, $p \leq .0001$ ($r_S = -.60$, $p \leq .001$). The reason for including words with a frequency of 9 or more for the Swedish sample was to have approximately the same number of data points in the correlation.

For the Swedish sample, there is also a similar relation between basic level actions and their subordinates with regard to frequency and ordinal position. Listed below are the Swedish basic level actions and their subordinates. For these groups, there is some corroborating evidence from Viberg (1992) for the basic level and subordinate relations between the locomotion verbs. Additional support was gained by informally asking native speakers to confirm the groupings. Approximate English translations follow the Swedish words. The TFs and MOPs are also presented in parentheses.

Springa (Running) (35:7.09) => jogga (jogging) (10:15.70), löpa (sprinting) (3:10.67), kuta (running energetically) (2:24), mila (running a 10k race) (1:29)

Gå (walking) (29:9.62) => promenera (strolling) (5:17.8), vandra (hiking) (2:9.5), lunka (walking at a leisurely pace, "moseying") (2:38), hasa (staggering about) (2:27), spatsera (sauntering) (1:14), marchera (marching) (1:15), flanera (strolling or wandering aimlessly) (1:6)

Hoppa (jumping) (32:5.31) => skutta (skipping or taking small hopping kind of steps) (2:7.5), hoppa rep (jumping rope) (1:3), hoppa häck (jumping hurdles) (1:4)

Prata (talking) (12:19.83) => tala (speaking or talking) (7:24.43), skrika (screaming) (13:24.92), viska (whispering) (3:29), ropa (shouting) (2:12), gnälla (complaining, whining) (1:30), argumentera (arguing) (1:16), diskutera (discussing) (1:27)

Some exceptions to the general trend that subordinates typically have a lower frequency and a higher ordinal position are found in the groups. Where this occurs, however, the frequency is very low and reflects the fact that one or two subjects accessed these words first. This is not the case though with PRATA. In that group, SKRIKA was listed by one more subject than PRATA, but the MOP was quite a bit higher. The gist of the data do tend to show, however, that the patterns for the group of 39 Swedish speaking subjects and the English group mentioned above are quite similar.

3.4.2.2 Swedish and American English Comparisons

In the following analyses, the stability of the action categories was determined on the basis of correlations between the two language samples. One problem with doing these analyses is the matching of the semantic similarity of the verbs for the two languages. Within the context of the Battig and Montague study, comparisons between the Maryland and Illinois samples could be done on the basis of the orthographic form of the words. For example, in the category of A FOUR-FOOTED ANIMAL, the occurrence of the word DEER among the lists for the Maryland subjects was assumed to have the same meaning as DEER in the Illinois sample. There was no question of any kind of regional difference in meaning between the two orthographically identical items. For the current analyses, there can be no such matching of items based on orthographic similarity. Instead, the words need to be matched according to their semantic similarity. Appendix B contains the list of the matched words and their respective TFs and MOPs. The list shows that 30 pairs of words could be closely matched according to their meanings. For example, the Swedish word SPRINGA has the same semantic content as the English word RUN. They refer to the same kind of pattern of bodily activity. There were, however, a number of words that did not match up quite so well. In these cases, the semantic domain of a word in one language was best matched by including the domains of two or more words from the other language. The semantic domain of PUSHING, for example, has no single Swedish equivalent. A group of four Swedish words was needed in order to match the semantic domain of PUSHING. As an example of the other kind of relationship, the Swedish word RIDA means to ride on an animal of some kind. It is not, however, used to refer to riding in a vehicle. Typically, it is used in the sense of HORSEBACK RIDING or RIDING A HORSE. The English word RIDING was included in the group because it can also mean HORSEBACK RIDING.

According to the first analysis, stability is a function of the degree of agreement between the TFs for the words that occurred 10 or more times in both lists. The correlation between the TFs for the two samples resulted in a coefficient of .64, $p \leq .0001$ ($r_S = .51, p \leq .01$). Comparing the ordinal positions of the verbs in this sample also revealed a significant correlation, $r = .50, p \leq .005$ ($r_S = .44, p \leq .01$). Discarding the 8 cases where more than one word was included in matching semantic domains revealed an improvement in the strength of the correlations, for the TFs $r = .70, p \leq .001$ ($r_S = .54, p \leq .01$) and for the MOPs $r = .61, p \leq .0005$ ($r_S = .56, p \leq .01$).

These results indicate a significant and rather large degree of stability across the two groups. This finding should be evaluated in the context of the Hampton and Gardiner (1983) findings mentioned above. Although none of the categories there dealt specifically with actions, the response frequency correlations of .64 and .70 are not much less than the mean of .76 for the correlations between the 12 categories used for the British and American groups. The stability found here, however, is not unequivocally robust. The change from one culture and language to another has a diminishing effect on the stability of the categories. As noted by Hampton and Gardiner (1983), ”[A]ssociative frequency may be expected to reflect local differences in language use and item familiarity.” It could, however, be argued that given this effect, there remains a relatively high degree of stability.

Multidimensional Scaling Analyses

The ordinal positions of the words on the subject’s lists can be viewed as indicating the semantic organization or association patterns between words. I should be clear about the fact that the mean ordinal position is not itself a distance measure. It is rather the case that the ordinal positions can be used to derive distances between the list items. As subjects think of words, the task can be seen as free association in the sense that words with similar meaning will tend to prime other words that share the same semantic domain. By going through the subjects' lists, the different ordinal distances between commonly shared verb pairs was determined and used to construct proximity matrices. The matrices were then used as input data in a multidimensional scaling program in order to get a multidimensional spatial interpretation of the ordinal structures inherent in the lists for the two groups.

The basis for obtaining proximity data based on ordinal position can be found in the work of Roger Shepard. According to Shepard (1962a, b), the structure of ordinal scale data is roughly isomorphic to metric axioms which allow the ordinal data to be monotonically transformed to an interval scale. Given this scale, the items can be given the interpretation of occurring in a psychological space in terms of a Euclidean metric. At this point it becomes meaningful to discuss the distances between various items in the space. On the basis of these distances, one can then talk about the dimensions that structure the space as well as information about the groupings that occur within it. It is in this sense that the proximity data obtained from the original lists can be used to reflect the psychological distances between the various items.

The notion of cross-cultural stability as defined as a function of the amount of agreement between the derived distances for the shared verb meanings in the two lists is much stronger than the notion of cross-cultural stability based on frequency alone. The correlations mentioned above only rely on a small subset of the possible pairs of words that the two lists have in common. MDS, on the other hand, provides distance measures between all possible pairs of words by taking their ordinal proximities into account. The main purpose of the following analyses is to determine the extent to which the English and Swedish groups agree with respect to the derived distances between semantically similar verb pairs. For example, the derived distances between RUN-WALK, RUN-JUMP, RUN-SWIM, WALK-JUMP, WALK-SWIM, JUMP-SWIM will be compared to the derived distances for their semantically similar Swedish counterparts. It is this much finer grained measure of cultural stability that is being tested for below, and to my knowledge, this method constitutes a novel approach to measuring such stability. For the correlations below, only the Pearson *r* will be reported since the distance measure represents a ratio scale.

The process of selecting the words to be included in the English and Swedish proximity matrices was simply done by taking all the verbs with a frequency of 8 or more from the English list and then taking the verbs with similar meaning from the Swedish list. One important limitation in selecting the English–Swedish verb pairs was that verbs with very low frequencies could not be included because this would result in too many missing values in the proximity matrices. I did, however, try to include as many common verbs as possible given the limitation. The resulting lists, 37 verbs each, are slightly different from the lists in Appendix B for the frequency correlations. (An asterisk next to the items in Appendix B indicates they were included in the MDS analyses.) The ordinal distances also indirectly reflect the influence of some of the other verbs not included here. This is due to the fact the ordinal positions of the verbs are a result of the intervening ordinal positions of verbs that are not included in this sample.

The next step was to construct a matrix for all possible pairings (37(37-1)/2 = 666) for the separate English and Swedish lists. For all such pairs and for all 39 subject lists, the absolute ordinal distance between the pairs was determined. Finally, the ordinal distances for all the pairs were averaged across subjects to obtain a mean ordinal distance for the item pairs according to formula 1.

$$(1) \quad \sum_{i=1}^{n} | d_i | / n ,$$

where d_i equals the absolute ordinal distance between a given verb pair, and n equals the number of times that a verb pair occurred across subjects. Due to the varying frequencies of the individual verbs, the frequencies of verb pairs also varied a lot. It should be noted that some of the verb pairs did not occur on any of the lists. These were left as missing data in the matrix. The number of missing values in the matrix for the Swedish verbs was 47, whereas in the English matrix, the number was 8. The resulting proximity matrices were then subjected to nonmetric multidimensional scaling. The statistical package SYSTAT™ was used to scale the data.

The first attempt at evaluating the cross cultural stability between the two sets of data was carried out by creating English and Swedish matricies for the 37 verbs that they had in common. MDS solutions in 3 dimensions were then produced and the resulting interpoint distances for all possible verb pairs were saved in separate files. The interpoint distances were used to determine the extent to which the respective semantic spaces for the two groups were similar. For example, the distance between the English pair RUN-JUMP was compared with the semantically equivalent Swedish pair SPRINGA-HOPPA. All 666 matched pairs were included in the correlation analysis. A Pearson r revealed no significant correlation between the two groups, r = .044. Based on the mean ordinal distances, there was no indication of cross cultural stability between the semantic spaces for the 37 matched verb pairs.

One reason for the lack of agreement may be that formula (1) is not an adequate measure of the association strength between the various verb pairs. Given the fact that many of the mean ordinal distances are based on relatively low frequencies (n) for the word pairs and are greatly influenced by extreme values (d_i), they tend to lead to unrepresentative rank orders in the data. For example, in the English data, the distance between the pair RUN-CHEW (8.75) is shorter than the distance between RUN-SWIM (11.00), where RUN-CHEW has a TF of 8 (SD = 4.59), and RUN-SWIM has a TF of 24 (SD = 9.55) The SDs indicate a large influence of extreme values for the pair RUN-SWIM. Whether or not this influence can be regarded as “undue” depends on the position one takes with respect to how much weight should be given to the two measures that reflect association strength, i.e., (1) the frequency with which a given verb pair occurs across the subjects’ lists and (2) the ordinal proximity between the two verbs. This would not be an issue, however, if frequency and ordinal distance were perfectly correlated, which they are not. Since the correlation between the two groups is stronger when TF is used than when MOP is used, I chose to reassess the stability between the groups by giving more weight to the frequencies as shown in formula (2).

$$(2) \quad \frac{\sum_{i=1}^{n} | d_i | / n}{n},$$

A second set of MDS analyses was carried out using the “weighted” mean ordinal distances between all possible pairs of the 37 verbs shared by the English and Swedish lists as described above. MDS solutions from 1 to 5 dimensions were obtained for the English and Swedish “weighted” proximity data. The stress and proportion of explained variance (R^2) values as a function of dimension for the English verb ordinal distance matrix are presented in Table 3.1.

As for the analyses based on the proximities using the “unweighted” mean ordinal distances, the 3-dimensional solutions obtained here will serves as the basis for the following analyses. The main reason for using the 3-dimensional solution has to do with the fact that correlations using distances obtained from 4-dimensional solutions did not improve the coefficients despite the reduced stress and increase in explained variance associated with the 4-dimensional solutions. In addition, a number of other starting configurations were tried without leading to any improvements in the values. I will also avoid interpretations about what qualities the dimensions might represent. This is due to the difficulty of interpreting the dimensions in a way that makes any sense with respect to the spatial solutions. For the interested reader, the 3-dimensional solutions for the English data are presented in Appendix C.

As in the first attempt at evaluating the general agreement between the 3-D solutions for the English and Swedish data, the derived distances for all possible verb pairs of the 37 shared verbs were subjected to a Pearson Product moment correlation. In contrast to the first attempt, a comparison of the derived distances based on the “weighted” means showed a significant correlation between the two groups, $r = .32$, $p \leq .0001$. This should be viewed in contrast to the results obtained for the proximity matrices based strictly on the “unweighted” mean ordinal distances where there was virtually no correlation present. The correlation, however, is not very strong. It only accounts for roughly 10 percent of the shared variance between the derived distances for the two groups. In addition to error in accounting for the remaining 90 percent of the variance, it is undoubtedly the case that there is a large amount of variation due to strictly cultural and individual differences. Within the context of the correlations above where TF and MOP were used in separate correlations, it is not so surprising that the much more sensitive measure used here results in a lower coefficient.

Table 3.1. Stress and R^2 as a function of number of dimensions for multidimensional scaling solutions for American English verbs.

	Dimensions				
	1	2	3	4	5
English					
Stress	.456	.293	.209	.147	.120
R^2	.36	.52	.64	.78	.82
Swedish					
Stress	.447	.283	.210	.157	.127
R^2	.39	.56	.64	.74	.78

According to the 3-D solutions for the English and Swedish proximity matrices, there was a tendency for two groups of verbs to cluster together. The first group has to do with motion/location verbs like RUN, JUMP, WALK, SWIM, JOG, etc. The second group is comprised of verbs that have to do with vocal or mouth actions such as TALK, LAUGH, CRY, SING, SCREAM, etc. For the motion/location verbs there is some corroborating evidence for the grouping of these verbs from two very different sources. Dittrich (1993) used biological motion displays of running, going up stairs, leaping, and jumping as exemplars of locomotory actions. The recognition of these biological motions was juxtaposed with the recognition of various social and instrumental actions. Relevant to the discussion here was the finding that subjects were significantly better at recognizing the group of locomotory actions than either the social or instrumental actions. The other source of evidence comes from Fisher, Gleitman, and Gleitman (1991) where they constructed verb triads that were presented to subjects in a similarity judgement task. The classes of verbs used in their experiments were selected in order to represent broad semantic distinctions, for example, perception/cognition verbs (look, see, listen), motion/location verbs (run, jump, throw, crawl, walk), and symmetrical verbs (meet, marry, match, join). On the basis of an overlapping cluster analysis, they found a distinct tendency for the motion/location verbs to be clustered together. For the group of vocal actions, I do not know of any other evidence that suggests such a grouping.

In addition to the trend towards a grouping of the vocal and motion verbs where they are relatively close to one another, a stronger test of their relation to one another in the context of cross cultural stability would be to see if their distances to one another are similar in the context of their respective distances to all the other words in the sample of 37. The notion here is that it is not just the distances between the motion and mouth verbs per se that is important here but their distances to one another as a result of their distances to all the other words. If the two samples differ with regard to the distances between the motion and vocal verbs and all the other verbs, then there will be little agreement between the Swedish motion and mouth verbs and their semantically similar English counterparts. While it may not be the case that the semantic spaces for the English and Swedish verbs agree more than is reflected by the correlation above, it may be the case that there are subgroups in the semantic spaces that are in greater agreement in the context of all the verbs.

In the following analysis, the motion verbs were chosen to represent actions that were largely restricted to bodily movement. The 9 bodily movement verbs selected for comparison between the English and Swedish solutions were: RUN, WALK, JUMP, SWIM, DANCE, SIT, JOG, STAND, and WAVE. The distances between all possible pairings of these verbs were taken from the list of the 666 distances for all pairings of the 37 verbs. A correlation between the English and Swedish distances revealed a significant and much stronger relation for this group of verbs than for the whole group of 37, $r = .66$, $p \leq .0001$. In contrast to this result, a correlation between the distances for a group of 9 randomly selected verbs from the group of 37 resulted in a much lower coefficient, $r = .29$, $p > .05$. The verbs included in this comparison were SWIM, TALK, SLEEP, CRY, SIT, JOG, HUG, ROLL, and DRINK.

A similar analysis of the verbs for vocal or mouth actions was conducted with the following 8 words: TALK, LAUGH, CRY, KISS, SMILE, SING, SCREAM, and SNEEZE. Somewhat surprising was the strength of the correlation for the interpoint distances between all possible pairing of these verbs, $r = .75$, $p \leq .0001$. This indicates that, as for the motion verbs, there is a large degree of cross cultural stability for the spatial solutions of these verbs given their distances to all of the other verbs. As a control, a group of 8 verbs was randomly selected, and the correlation between the interpoint distances for this groups resulted in a coefficient of .31, $p > .10$. The verbs in this sample were: SWIM, WRITE, THROW, EAT, CRY, HUG, STAND, and PAINT.

In characterizing the results of the correlations based on the interpoint distances between the verbs, there remains a significant amount of cross cultural stability between the commonly shared verbs included in the analyses when frequency is given more weight than the ordinal configuration of the verbs. Although the magnitude of the correlation is weaker than when TF and MOP were tested separately, this should be viewed in the context of the greater sensitivity of the MDS based analyses. The correlations involving the two subgroups point towards a more substantial stability for the verbs in those groups. And to the extent that these verbs, with the exception of JOG perhaps, can be considered basic level concepts, they appear to be similarly organized in the network of commonly shared verbs between the Swedish and English groups.

3.5 General Discussion and Conclusions

First of all, the findings presented here should be viewed in the context of an important assumption that stems from previous work in categorization. This assumption concerns the hierarchical relation among concepts. Given that this hierarchical relation exists, one can reasonably assume that it can be accessed by subjects when presented with an appropriate task. In contrast to verification and identification tasks where subjects categorize stimuli in relation to a given level in a concept hierarchy, the task used here assessed the hierarchical relation between general perceptual criteria as applied to actions and its subsumption categories in a top-down fashion. The general criteria can be seen as denoting a very general superordinate for a wide range of actions, and that when presented with the free listing task, subjects had no difficulty of generating action verbs in relation to it. In a sense similar to the findings of Rosch, Mervis et al. (1976) for objects, actions have a visual shape. This shape is largely formed by invariants having to do with the pattern of movement of body parts rather than social context and goals. This is not to say that these factors play no role in action categorization, but rather that there is a domain of action categories for which perceptual criteria can reasonably be applied and can serve as a basis from which to list those actions without apparent conflict with social context and goals.

The lists of verbs provided two basic measures- response frequency and mean ordinal position- that were used to investigate which verbs indicate a basic level for actions and the degree to which the verbs were stable across Swedish and English. Within the context of the British-American frequency correlations for object categories, the Swedish-American correlations reflect a significant amount of stability. Recall that the mean correlation between the 12 object categories used in the Hampton and Gardiner (1983) study and the same 12 categories from Battig and Montague (1969) resulted in a coefficient of .76 for the frequencies. Given the fact that the correlations presented here deal with action categories and that the superordinate description by which subjects generated their lists was more general than the superordinates used in Hampton and Gardiner, the coefficient of .64 for the verb correlations, including semantic domains composed of more than one verb, is strikingly high. The reason for comparing the results here with the Hampton and Gardiner results is that it appears to be the only such correlation available that uses response frequency as a measure of cross cultural stability.

If doubts still loom as to the degree of stability between the most frequent verbs in the English and Swedish samples, the more sensitive measure of the interpoint distances for both groups provides additional evidence in its favor. Although the overall correlation for the 666 interpoint distances for the 37 verbs was relatively small, .32, the correlations for the motion and vocal actions were much larger, .66 and .75 respectively. My claim here is that this is evidence of a shared semantic organization between English and Swedish subjects for these verbs in relation to the other verbs that were included in the MDS analyses.

Concerning the issue of the hierarchical status of the actions included in the analyses presented above, there is also evidence to support the claim of an existing basic level=>subordinate level relation for actions. If basic level action categories can be accessed via perceptual criteria and are usually accessed first in contrast to subordinate level actions, then this relation should reveal itself when using a free listing task, for example. This was shown to be the case for both English and Swedish lists. Whereas verbs like RUN, JUMP, WALK, and TALK were listed often and early on, their respective subordinates were generally listed less often and later on in the subjects' lists. These results mirrored similar basic level=>subordinate relations in the Battig and Montague data.

For the verb results, it is unlikely that this effect is an artifact due to the nature of the instructions in the sense that they favored basic level verbs over subordinate level verbs by emphasizing perceptual criteria. It does not seem to be the case that subordinate level actions are less "perceptual" than basic level actions. MARCHING, as a subordinate to WALKING, for example, has distinct visual properties in the way in which one's legs and arms move in relation to one another. If this is the case, then one is led to the further question of the relation between "perceptual shape" and its role in determining the basic level. In other words, why are MARCHING and TIPTOEING subordinate and WALKING basic level? Suffice it to say here that "perceptual shape" is most certainly implicated in the formation of many concept hierarchies, and, as such, it is a valid variable to manipulate or measure when investigating basic level effects. In the context of the claims made here, the perceptual criteria were not manipulated or measured, but used to get subjects to list ANY actions that seemed to match the general description. The explanation for WHY the lists were structured the way they were has to do with the ways in which semantic memory and the lexicon are structured with regard to, for example, "category utility" (Corter & Gluck, 1992) and the linguistic community. The perceptual criteria were simply intended to induce the subjects to "dump" a portion of their semantic memory onto a sheet of paper.

Finally, I should clarify the claims that I am not making. First, I am not claiming a cross cultural stability for all basic level concepts. And I am not making the weaker claim that all basic level concepts are inherently stable between individuals within a culture. First of all, there is a lack of uniformity between the individual lists for both Swedish and English groups, as shown by the distribution of the frequencies. Secondly, out of all the verbs listed by the subjects, only a small portion of them were included in the stability analyses. Indeed, there are major differences between individuals and cultures that have their explanations in numerous factors, which I will not go into here. The claim that I do want to make concerns the issue of whether there are any factors according to which these differences are minimized. To this issue, I want to reply, YES, with the reservation that the domains for which the differences are minimized are possibly rather few, at least in the context of action categories. Being able to say, however, precisely WHAT those factors are and HOW they minimize the differences is much more difficult than simply saying that they exist on the basis of the results. Although I will refrain from a precise description of the factors, I will offer a general suggestion as to what they might be.

Given the assertion that the two cultures discussed here are very close to one another, it is not so surprising that verbs like WRITING, READING, DRIVING, and, to certain extent, BIKING occur frequently in the lists. This may also be due to the fact that college students served as subjects. But among a large portion of the remaining verbs that were used in the MDS analyses, one sees verbs denoting actions that have to do with fundamental ways in which people move, and interact with their surroundings, as in KICKING, THROWING, EATING, HITTING, LIFTING, DRINKING for examples of the latter. Along a similar line, the vocal and facial actions indicate basic ways of communicating and expressing emotion. The one exception here is SNEEZING, which is yet a distinct action that involves a very salient "visual shape."

The two factors that I propose as playing a significant role in minimizing cultural and individual differences are (1) the physical constraints (invariants) involved in human motion, and (2) the frequency with which people engage in certain actions or see other people performing them. This proposal is not new in the sense that physical (dynamic) invariants have been suggested as an explanation as to why people are so good at recognizing actions in point-light displays and that "frequency of instantiation" plays a strong role in predicting graded structure in categories (Barsalou, 1985). The relation between the factors is one of unidirectional dependency where the notion of physical invariants is the more primitive of the two. Whereas it seems reasonable that physical constraints would have a bearing on how frequently an action occurs, the opposite dependency does not seem plausible.

As to the issue of HOW the above mentioned factors minimize the differences, one has to consider the extent to which cultural and individual circumstances, i.e., context, can limit the occurrence of a given action. For example, while KICKING and THROWING may be limited by the extent to which various sporting activities involve these actions, RUNNING, WALKING, TALKING, etc. are not likely limited to the same extent. Physical invariants in human action are quite constant across individuals and cultures. These invariants and frequencies of actions minimize the differences by determining the range of actions that can be performed in conjunction with the cultural and individual limitations that affect the frequencies of the range of actions.

Insofar as certain actions are fundamental to human activity and face few cultural limitations, I propose there will be a strong tendency for this to be reflected in the lexicon. While the range of actions that can be performed is nearly infinite, cultural and linguistic communities place limits on both frequencies and “how an action shall be called” (cf., Brown, 1958).

APPENDIX A

Total Frequency (TF) and Mean Ordinal Position (MOP) for American English Action Verbs and Phrases.

Item	TF	MOP
running	115	4.37
walking	99	7.32
jumping	92	7.61
swimming	65	13.94
skipping	61	7.54
writing	56	20.45
talking	56	13.71
eating	55	17.60
sleeping	54	17.03
throwing	53	16.43
sitting	48	15.81
kicking	47	15.45
hitting	46	15.63
crying	46	18.35
dancing	46	19.96
laughing	44	16.96
smiling	43	19.49
standing	40	14.48
jogging	39	11.87
driving	38	19.05
yelling	38	17.50
falling	37	18.89
blinking	35	15.20
pushing	34	21.00
lifting	33	19.61
drinking	32	21.00
hopping	32	8.47
kissing	30	22.60
singing	30	17.43
pulling	30	22.33
reading	28	22.75
catching	27	21.78
waving	27	18.11
scratching	26	20.35
touching	25	15.36
hugging	25	25.88
playing	25	19.04
punching	24	17.67
screaming	24	21.21
sneezing	24	15.71
coughing	23	19.04
rolling	22	22.46
bending	22	18.27
stretching	22	19.59
diving	22	22.46
frowning	22	22.50
climbing	21	19.24
painting	21	32.48
dressing	21	23.38
biking	21	15.81
chewing	20	15.85
crawling	20	16.65
shaking	20	22.15
drawing	19	28.47
skiing	19	19.37
moving	19	16.21
swinging	19	20.11
exercising	18	18.22
leaping	18	18.67
looking	17	25.82
flying	17	15.41
carrying	17	22.00
whispering	17	24.94
tripping	16	19.44
fighting	15	15.73
cleaning	15	29.67
watching	15	24.53
twisting	15	23.47
speaking	14	19.50
cooking	14	24.64
spinning	14	26.21
brushing teeth	13	16.46
holding	13	22.62
brushing	13	24.69
staring	13	29.46
winking	13	22.31
wiggling	13	22.08
laying	12	23.67
grabbing	12	22.00
washing	12	33.58
slapping	12	23.50
showering	12	26.25
kneeling	12	22.08
studying	12	22.42
snoring	11	28.73
flexing	11	20.09
spitting	11	21.27
leaning	11	22.73
shouting	11	25.00
reaching	11	20.46
clapping	11	17.09
rubbing	11	26.82
turning	11	17.55
closing	11	29.27
opening	11	29.73
smelling	10	19.00
seeing	10	13.60
squatting	10	21.40
galloping	10	11.00
acting	10	19.50
riding	10	23.60
riding a bike	10	24.20
tying shoes	9	26.11
sprinting	9	13.89
working	9	23.78
brushing hair	9	18.00
sliding	9	21.89
shooting	9	26.56
squeezing	9	24.89
tapping	9	18.89
biting	9	19.11
twitching	9	20.11
bouncing	9	16.11
yawning	9	18.78
grasping	9	19.22
poking	9	24.11
hiking	9	19.11
taking	9	25.00
itching	9	14.55
giving	8	27.25
whistling	8	21.50
fishing	8	19.25
nodding	8	15.75
shopping	8	25.38
cutting	8	30.63
typing	8	27.38
pointing	8	14.13
tossing	8	25.88
standing up	8	12.50
shoving	8	23.75
squinting	8	14.88
breaking	8	35.50
skating	8	15.63
hearing	7	14.43
killing	7	19.57
breathing	7	10.14
pinching	7	22.57
swallowing	7	27.71
sniffing	7	19.14
dropping	7	21.86
sitting down	7	12.71
bending over	6	18.83
twirling	6	32.50
stopping	6	19.00

APPENDIX A (continued)

sculpting	6	47.17
pounding	6	39.50
sledding	6	28.67
bathing	6	23.00
snapping	6	18.83
sweating	6	17.67
giggling	6	15.67
feeling	6	23.33
snapping-fingers	6	17.00
smoking	6	19.00
stumbling	6	17.83
resting	6	25.17
driving a car	6	23.17
dribbling	6	19.67
urinating	6	18.33
flipping	6	17.67
shivering	6	19.33
listening	5	19.80
grimacing	5	31.00
playing an instrument	5	25.00
petting	5	26.40
trotting	5	16.20
comb hair	5	23.60
tasting	5	18.00
choking	5	19.80
riding a horse	5	22.60
strolling	5	17.00
calling	5	33.80
sewing	5	37.80
stepping	5	17.40
laying down	5	27.00
slipping	5	33.20
destroying	5	31.80
stomping	5	23.60
licking	5	12.40
blowing	4	24.50
sucking	4	21.50
relaxing	4	24.00
pacing	4	25.00
crossing legs	4	17.25
lying down	4	19.50
going to the bathroom	4	24.75
shaving	4	34.25
stealing	4	43.00
tumbling	4	18.75
helping	4	32.50
setting	4	36.75
scribbling	4	40.50
combing	4	39.25
sweeping	4	31.00
coloring	4	33.00
shaking head	4	20.75
pouting	4	27.75
caressing	4	13.25
making love	4	19.50
praying	4	26.50
picking up	4	24.50
extending	4	23.00
scraping	4	24.25
burping	4	29.25
buying	4	30.00
cracking	4	24.75
rowing	4	19.50
hiding	4	35.00
puking	4	23.75
tapping foot	4	20.00
digging	4	28.50
asking	4	23.75
swaying	4	25.75
beating	4	36.25
washing face	4	17.75
throwing a ball	3	9.33
thinking	3	15.33
having sex	3	12.33
wiggling toes	3	23.67
blowing nose	3	26.67
picking nose	3	18.33
putting	3	29.00
hammering	3	39.00
surfing	3	39.67
smirking	3	24.00
juggling	3	31.67
making a fist	3	26.33
lie down	3	24.67
cracking-knuckles	3	18.67
stabbing	3	29.00
contracting	3	26.00
vacuuming	3	29.33
sketching	3	26.67
teaching	3	43.67
spiking	3	32.33
rotating	3	25.00
gazing	3	34.67
wiggling-fingers	3	20.67
raising arm(s)	3	8.00
squirming	3	14.00
sighing	3	27.67
flicking	3	18.00
sitting up	3	28.67
pasting	3	34.33
baking	3	38.67
sailing	3	31.00
scrubbing	3	34.33
turning around	3	23.00
waking up	3	23.00
grinning	3	23.33
pouring	3	29.67
banging	3	36.67
picking	3	18.00
smacking	3	21.33
puckering	3	32.00
opening door	3	21.33
farting	3	26.00
counting	3	27.33
attacking	3	17.00
lifting weights	3	24.67
vomiting	3	13.67
limping	3	32.00
lie	3	12.67
hobbling	3	18.67
closing eyes	3	22.33
fidgeting	3	27.33
bowing	3	23.33
going	3	22.00
training	3	33.33
receiving	3	34.33
passing	3	24.67
shuffling	3	19.67
doing jumping jacks	3	16.33
undressing	3	32.00
nodding your head	3	13.67
crouching	3	20.33
paddling	3	32.00
using	3	28.67
hurting	3	25.00
selling	3	24.00
tearing	3	14.67
screwing	3	14.00
doing a cartwheel	3	11.33
playing sports	3	17.33
somersaulting	3	29.00

APPENDIX B

VERB Pairs: English – Swedish

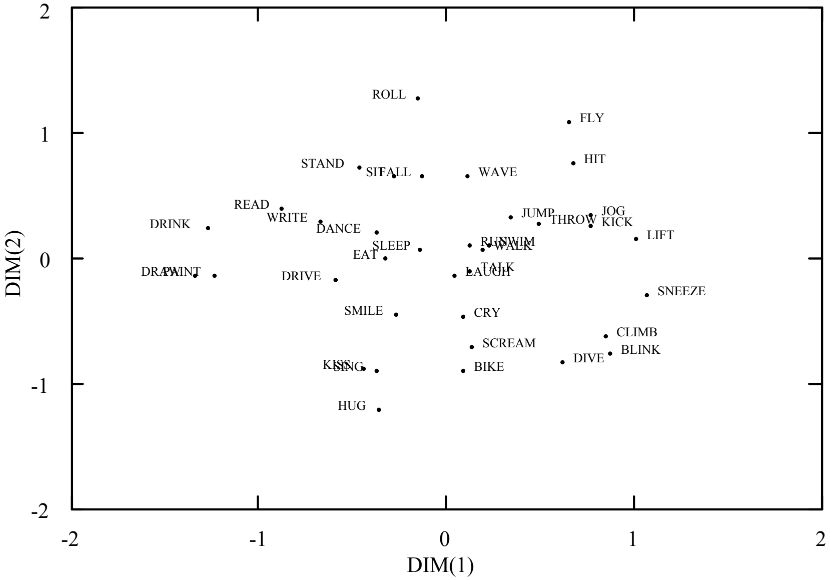
Item – English	**TF**	**MOP**	**Item – Swedish**	**TF**	**MOP**
1. running*	38	4.08	springa*	35	7.04
2. walking*	30	6.83	gå*	29	9.62
3. jumping*	27	7.48	hoppa*	32	5.31
4. hopping	10	11.60	skutta	2	7.50
5. swimming*	24	12.42	simma*	23	13.04
6. talking*	22	12.14	prata*	19	22.13
			tala		
7. writing*	21	21.95	skriva*	29	15.14
8. sleeping*	20	18.05	sova*	16	15.06
9. throwing*	19	12.84	kasta*	7	19.86
10. eating*	19	17.68	äta*	24	13.17
11. laughing*	17	14.94	skratta*	20	18.65
12. dancing*	17	21.65	dansa*	18	17.28
13. crying*	17	17.65	gråta*	19	23.11
14. kicking*	16	14.44	sparka*	4	24.50
15. falling*	15	21.00	falla	7	17.50
			trilla*		
16. pushing	14	22.00	putta	10	28.67
			knuffa(s)		
			köra		
			trycka		
17. sitting*	14	19.14	sitta*	17	18.88
18. kissing*	13	27.23	pussas*	18	25.50
			kyssa(s)		
19. hitting*					
punching	24	15.18	slå*	15	22.60
20. smiling*	13	17.77	le*	9	20.56
21. lifting*	12	12.67	lyfta*	7	23.86
22. jogging*	12	9.33	jogga*	10	15.70
23. driving*	12	19.25	köra	16	19.84
			köra bil*		
24. pulling	11	24.64	draga	4	33.75
25. screaming*	9	20.44	skrika*	13	24.92

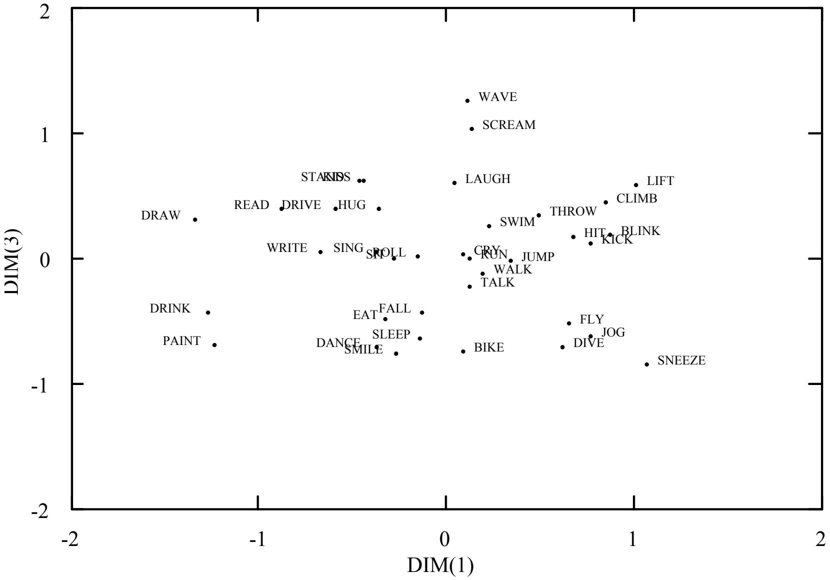
APPENDIX B (continued)

26. reading*	11	19.82	läsa*	13	15.23
27. hugging*	10	29.30	krama(s)*	15	24.33
28. climbing*	10	20.60	klättra*	6	24.67
29. standing*	10	17.50	stå*	9	18.00
30. rolling*	10	23.90	rulla*	5	26.80
31. singing*	10	19.10	sjunga*	14	22.86
32. catching	10	18.60	fånga	1	10.00
33. riding a bike					
biking*	11	23.79	cykla*	24	9.83
34. drinking*	9	28.44	dricka*	15	19.13
35. making love	1	27.00	älska	21	19.10
36. painting*	8	30.75	måla*	13	25.69
37. drawing*	8	26.13	rita*	10	17.60
38. riding a horse					
horseback riding					
riding	4	22.00	rida	10	11.20
* = Words included in the MDS analyses					
Additional words included i the MDS analyses					
flying	9	19.33	flyga	6	19.17
blinking	8	13.63	blinka	7	21.00
sneezing	8	14.63	nysa	4	30.00
waving	8	20.75	vinka	8	19.75
diving	8	18.13	dyka	9	17.44

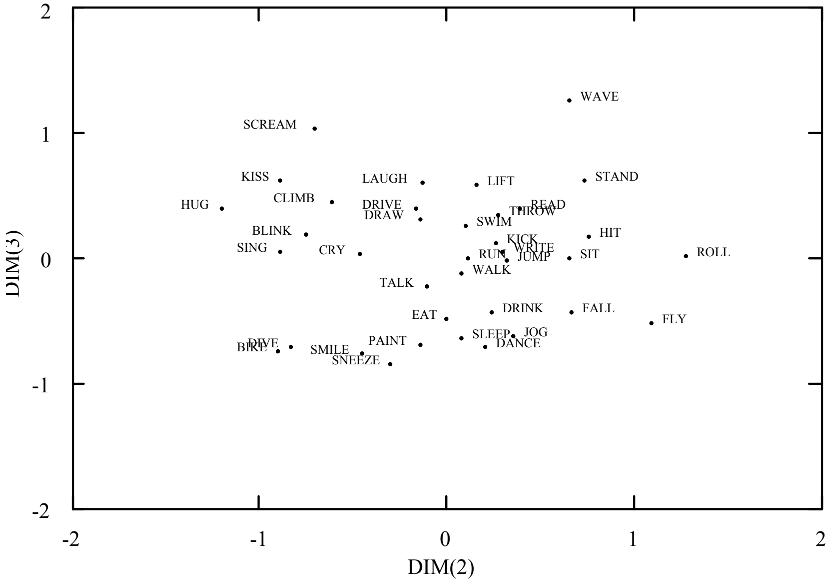
APPENDIX C

Three-dimensional MDS solutions based on the ordinal proximities for 37 English verbs.





APPENDIX C (continued)



Chapter 4 - Biological Motion and The Point-Light Technique[28]

4.1 Overview

This chapter presents a brief review of the development of the point-light technique for creating displays of human biological motion. I will also present the method used to construct the biological motion displays used in the experiments presented in chapters 5 and 7. Although the technique of creating the stimuli for the different experiments varied somewhat, the basic method was the same. A further purpose of the chapter is to present the technique as a useful tool for studying action perception.

The biological motion stimuli used in the experiments presented in chapters 5 and 7 were created in the early 90s. At that time, available computer hardware and software put severe limitations on the use of techniques that are now much more developed and less expensive. Despite this limitation, the technique used to create the stimuli still has its advantages in the context of current technological developments. These advantages will be addressed in this chapter.

When viewing common actions in natural settings, objects used in actions as well as the surroundings (physical and social) in which the actions take place provide a viewer with cues about the actions. In other words, the context can act as an effective constraint for recognizing the actions of others. The major advantage of the point-light technique is that it isolates the motion cues of the action from the contextual factors, since the latter are not visible. This, however, does not mean that contextual factors cannot be ‘seen’ by the viewer. This apparent contradiction can be explained by the fact that some contextual factors can act to physically constrain the spatiotemporal properties of human motion. To the extent this occurs, a viewer may be able to see this in the flow patterns of the patches or lights. For example, the weight of a lifted box places biomechanical constraints on how a person lifts the box (Runeson &

---

[28] Section 4.4 was presented at the following workshop: Primacy of Action: An advanced inter-disciplinary workshop, Manchester, England, 1993.

Frykholm, 1981, 1983). Further examples of this are walking in deep snow, or on ice where the structure of the supporting surface for walking constrains how a person traverses such surfaces.

The point-light technique represents a very useful tool for removing the explicit form cues of the human body as well as the objects that may be used in carrying out a specific action. The point-light technique allows researchers to isolate the dynamic properties of the human body by removing all other cues (e.g., explicit form, color, texture, etc.) that may guide action perception. This does not mean that form is not perceived or that form processing is not involved in action perception (Beintema & Lappe, 2001).

The point-light technique has also been applied to emotion perception (e.g., Atkinson, Dittrich, Gemmell & Young, 2004; Dittrich, Trocianko, Lea & Morgan, 1996). Other animals are also sensitive to the coherent pattern of motion of other individuals depicted in point-light displays. For example, cortical cells in non-human primates are sensitive to the direction of articulation in a point-light walker (Puce & Perrett, 2003 for a review). Blake (1993) has also shown that cats reliably discriminate point-light displays of cat motion from different kinds of motion-based foils including phase-scrambled motion. This ability, however, was not restricted to conspecifics. The cats were also able to discriminate a human point-light walker from a scrambled foil. Pigeons are also sensitive to point-light displays of pecking and walking (Dittrich, Lea, Barrett & Gurr, 1998).

4.2 Techniques for Creating Displays of Biological Motion

There are a number of currently available techniques for creating point-light displays of biological motion. For a review of the techniques, see Dekeyser, Verfaillie and Vanrie (2002). (See also Thornton (2006) for a more in-depth review of the point-light technique and its role as a research tool in the area of biological motion processing.) Before presenting the technique used in the experiments for this book, I will briefly describe two broad approaches to constructing point-light stimuli. The purpose is to place the technique that I have used within the context of currently available techniques.

Since Johansson (1973) first used the point-light technique as a research tool to demonstrate the visual salience of human action based on the perceptual grouping of the motion of 13 moving dots, researchers have performed many experiments to investigate this incredible ability of the visual system. Thornton (2006) estimates that over 500 published articles have been inspired by the Johansson point-light technique. The phenomenon of biological motion perception has spawned a research field that seeks to understand how the visual system is able to see meaningful motion (actions) in the mere movement of 13 dots attached to the joints and head of a human body.

4.2.1 Video-Based Techniques

Johansson (1973) originally discussed two different techniques for creating point-light displays. The first method used small flashlight bulbs attached to the joints of an actor dressed in tight fitting clothing. In a darkened room, the actor performed a number of actions, which were recorded using a 16 mm film camera. The apparent disadvantage of this method was the cumbersome attachment of the power supply to the flashlight bulbs. With more modern motion capture techniques, this is no longer a problem. Johansson (1973) employed a second technique where he used reflective material attached to the joints. This allowed the actor to move more freely and naturally. When a light source was directed towards the actor wearing the reflective patches, the motion of the patches could be isolated by adjusting the contrast and brightness controls on a video monitor. I will refer to this technique as "patch-light" instead of "point-light."

There are some (important) differences between the patch-light and point-light displays. Patches can potentially convey *form* information if they cover enough area around the joints of a human actor. They can also convey information about the direction of the light source. The extent to which people may actually use this information in the visual processing of patch-light displays has not been systematically investigated. It is, however, possible that the additional form information facilitates the recognition of a human figure. Therefore, if action recognition of patch-light displays relies on the visual processing of human form as well as motion information, patch-light displays may facilitate recognition to some extent.

The use of small light sources (LEDs or reflective material) while recording often leads to artifacts of occlusion. For example, if the orientation of the actor or a body limb changes while filming, the light source will not show up in the recording. In order to remedy this occlusion artifact when using reflective material, it is necessary to wrap the material around the joints of the body. The remaining occlusion effects will therefore be due to limbs passing in front of other limbs.

Depending on their size, the LEDs can become occluded by simple rotation of the axis of a local limb. For example, if an LED is placed on the wrist adjacent to the top of the hand, a rotation of the lower arm can lead to a disappearance of the LED despite the fact that the wrist is still visible to an observer. This kind of 'unnatural occlusion' should be distinguished from occlusion effects arising from whole body in-depth rotation and occlusion resulting from limbs passing in front of other limbs in naturally defined movement.

The use of video recordings has a number of advantages and disadvantages. One significant advantage is the cost of the technique. With a standard video camera and access to standard video editing software, patch-light displays of biological motion can easily be created and displayed. A further advantage is that the range of actions is not restricted to recording actions within a limited area (cf. motion capture techniques). The front-end time it takes to create the displays is also relatively short in comparison to other techniques that demand a more extensive investment in set-up time.

A major limitation of direct video recording, however, has to do with the ability to manipulate the individual patches, or groups of patches in order to investigate, e.g., various spatial parameters involved in the visual processing of biological motion. The key difficulty is in isolating or extracting the motion vectors of the patches/points in order to exert control over their placement in a display. There is, however, a method for overcoming this disadvantage to some extent. Photo editing software can be used to select the patches and then delete all other information. The patches can then be individually manipulated to create different spatial configurations such as creating dynamic masking elements. The disadvantage of this method, however, is that it can be quite time consuming if there is no alternative to frame-by-frame editing. A further limitation is that the display is 2D. In order to obtain displays of different orientations, the same action needs to be recorded again from a different orientation or multiple cameras need to be used when an action is performed.

The relative advantages and disadvantages of direct video recording for creating displays of biological motion depend on the specific issues being investigated. For example, Runeson and Frykholm (1981) used the patch-light technique in their study of estimations of the weights of lifted boxes. Dittrich (1993) investigated action identification, and Aktinson et al. (2004) explored emotion perception. Recently, Loula, Prasad, Harber and Shiffrar (2005) used this technique to address potential differences when we view our own actions compared to our ability to recognize the same actions performed by other individuals.

A variation of the direct video recording technique involves using video recordings of actions in "natural" settings or pre-existing filmed sequences of human movement, i.e., sporting events, children playing, etc. By using standard video editing capabilities available on many standard computers, it is possible to manually overlay point-lights on the joints of any moving creature (or object) and then simply save the point-light files as a video sequence (Mather & West, 1993). The obvious advantage to this technique is that it is possible to create displays of biological motion using actions from natural settings. It is also possible to extract the 2D coordinates of the points to create files that only contain information about the spatial coordinates of the patches (Grossman & Blake, 2002). A major drawback, however, is that this technique requires frame-by-frame placement of points to the original video sequence. A further consequence is that it is difficult to assign the point to the exact same joint position for every frame, which can lead to local jitter in the displays. There are, however, methods for smoothing the trajectories (Giese & Lappe, 2002).

4.2.2 Motion Capture

One major general advantage of the point-light technique in studying action perception is the ability to isolate motion information from information about object form as well as contextual information about the physical and social surroundings in which actions are performed. However, once the motion information is extracted, its systematic use in experiments of action or biological motion perception is constrained by the recording technique. As previously mentioned, video-based recordings typically lack the flexibility of directly manipulating the spatial coordinates of the markers, especially in 3 dimensions.

In contrast to video-based recordings, motion capture techniques allow an experimenter to work directly with the 3D coordinates of the point-light markers. The basic technique of motion capture systems relies on cameras or other kinds of sensors that detect the signals transmitted from markers placed on the body of a human actor. The number of sensors as well as their placement in a recording space can vary somewhat. Unlike Johansson's (1973) technique where flashlight bulbs were connected to a power source via wires, movement of the actor using a modern motion capture system is not constrained by wires attached to the markers placed on the actor. The *range* of movements, however, is constrained by the 3D recording space in which the recording sensors are arranged. Actions that require large spaces such as ice skating, skiing, swinging, climbing, etc. therefore can be difficult to capture.

Once the motion coordinates of the markers are captured and stored, they can be temporally and spatially manipulated to create different viewpoints, temporal variations as well as combinations of the two. For example, manipulations of temporal (e.g., Cutting et al., 1988; Mather et al., 1992; Thornton et al., 1998) and spatial variables (e.g., Bertenthal & Pinto, 1994; Dittrich, 1993; Mather, Radford &West, 1992; Thornton, Rensink & Shiffrar, 2002) reveal the importance of these factors for, e.g., judgments of identity, judgments of direction of articulation and figure detection. For specific details regarding these findings, see Thornton (2006). One disadvantage of this flexibility of motion capture systems is that software must be developed to make use of the information. Some routines for manipulating the 3D coordinates may be available in prepackaged animation software, but some customization may be needed depending on the specific manipulations needed for an experiment.

Although motion capture systems capture the naturalness and subtleties of human movement and allow for greater flexibility in systematically manipulating spatiotemporal parameters in displays of biological motion, there are some drawbacks. For example, the issue of occlusion arises in two ways. Firstly, occlusion artifacts can arise due to the fact that body and limb rotations block the reflected light from some markers. The result of this occlusion leads to missing coordinate data for some aspects of human movement. In this case, camera angles need to be adjusted to the specific movements so as not to loose too much data. A second occlusion issue is the recording of marker coordinates in 3 dimensions, which leads to coordinate data that *includes* marker coordinates for naturally occurring occlusions. In other words, when the 3D coordinates of the markers are played back, the point-lights associated with natural occlusions will still be visible. Dekeyser et al. (2002) have, however, developed a method to rectify this problem.

An additional potential drawback of motion capture systems is their cost, which can run into the tens of thousands of dollars. Motion capture systems typically require dedicated lab space as well. In short, the initial set-up time for motion capture systems can be quite extensive. Once the motion capture system is up and running, additional time is needed in order to learn and/or customize software applications that create and display the point-light stimuli.

The techniques of direct video recording using reflective patches and motion capture systems require an actor to wear markers of some kind. Since the recorded actions are produced for the purpose of recording action stimuli, they may not be entirely representative of human movement in natural settings. The ultimate goal of studying action perception is the ability to systematically investigate various factors that influence human movement as well as action recognition in natural settings. In this regard, development of the technique for recording human movement should include techniques for manipulating video recordings of actions that occur in natural settings. Thornton (2006) describes such a technique for studying gender recognition.

As the study of action perception develops, it will likely be the case that a greater emphasis may be placed on the social factors involved. This may create difficulties in using any motion capture tools that restrict human movement as well as the interaction between humans. It may also be the case that we might want to extract motion information from subjects who do not know that their movements are being filmed. The major point here is to suggest that action perception should move towards more naturalistic situations where unobtrusive techniques are needed to isolate human movement.

One disadvantage of the patch-light technique has been discussed by Berry, Kean, Misovich and Baron (1991) regarding the role of motion in the area of social perception. Since the patch-light displays involve wearing the appropriate patches and, hence, people are aware of being filmed, this may have an unwanted effect on the subsequent social interaction between individuals or groups. A less intrusive method of capturing the motion in social interaction proposed by Berry et al. is to video film social interaction without any patches and then use the method of quantizing the video sequences. This method effectively reduces the structural information in the sequences while preserving the inherent motion, and the experimental results are similar to those of the patch-light technique. Given the concerns raised in their article, the quantizing technique has the advantage of capturing motion from social interaction where the actors can be unknowingly filmed.

Currently, there are two developments that emphasize greater flexibility in creating and manipulating human motion stimuli. The first has to do with using motion capture data to animate solid-body models, in order to more fully understand the interaction between form and motion cues. The other development has to do with creating editing routines and algorithms for video recordings taken from naturalistic situations.

As a technique for isolating motion information from form information of the human body and the surrounding environment, video recording remains a cost effective and experimentally valid alternative to other techniques. If one, however, needs to spatially manipulate the point lights or manipulate their appearance (Ahlström, Blake & Ahlström, 1997), then video recording will not be appropriate.

4.3 Comparison Between Point-light Displays and Whole Body Motion

One reason for the interest in point-light displays of biological motion has to do with the fact that people (and other animals) can see the actions of others represented by the coherent motion of just 13 dots on a computer screen. In evaluating the usefulness of point-light display for studying action recognition, the question arises as to what extent action recognition performance differs for point-light displays and whole body motion. One might expect a large difference between point-light displays and whole body motion. In the latter case, there is an abundance of information about human and object form. If static form cues in point-light displays are drastically reduced, how does behavioral performance with point-light displays differ from displays where the whole body is present? A related question concerns differences in processing as revealed by neuroimaging studies. One difficulty in addressing potential performance differences between point-light and whole body displays is the use of different tasks and procedures in different studies.

Results from Grossman and Blake (2002) show for example that activation levels for point-light displays and whole bodies does not differ in the posterior superior temporal sulcus (pSTS). Both whole bodies and the 12 points of light were equally effective in activating pSTS. In this study, a 1-back task was used. Subjects were instructed to push a button whenever there were sequential repetitions of an action, a kind of matching task. Grossman and Blake (2002) used a number of different point-light actions, e.g., running, kicking, jumping and throwing. In addition to actions, subjects also viewed stationary images of bodies, faces and objects.

The results from Beauchamp, Lee, Haxby and Martin (2003) showed, however, somewhat different results regarding the activation of pSTS when viewing human videos and human point-light displays. Videos of fully illuminated human bodies elicited greater activation in pSTS than did human point-light displays. The task used in this study was different from the one used in Grossman and Blake (2002). Beauchamp et al. presented subjects with videos of actions (jumping jack, stair climb, jogging, soccer kick, etc.) and moving tools. The task was a two-alternative forced-choice task where subjects decided if the stimulus contained a human or a tool.

Using behavioral measures, Thomas and Jordan (2001) obtained performance differences in audiovisual speech perception for fully illuminated video displays and point-light displays. While point-light displays produced sufficient cues for audiovisual speech recognition, viewing fully illuminated faces led to significantly better recognition. In contrast to speech perception, Hill, Jinno and Johnston (2003) explored differences in point-light and fully illuminated facial motion when subjects were given a sex-judgment task. The fully illuminated facial motion was created by a motion capture system and then using the coordinates to animate an average face. Hill et al. found similar levels of performance for full view solid body faces and point-light faces.

Dittrich et al. (1996) showed that subjects were sensitive to the emotion conveyed in point-light displays of human dance, indicating that motion cues alone are sufficient to convey emotion in dance. Dittrich et al. also found a difference between fully illuminated dance and point-light displays. While subjects could reliably identify different emotions in point-light displays, emotion identification with fully illuminated dancers was significantly better. Aktinson et al. (2004) further addressed the issue of emotion identification in point-light displays and fully illuminated displays in a series of studies and obtained results generally consistent with previous findings from Dittrich et al. (1996).

The previously mentioned work of Giese and Poggio (2003) also showed that a simulation of their model demonstrates generalization from the recognition of full-body action to the recognition of actions (walking) depicted as point-light figures. Once the system is trained on full-body motion, it is also sensitive to point-light displays. This was the case for processing in the motion pathway, but not for processing in the form pathway.

4.4 Creating Patch-light Displays of Biological Motion

This section describes the chosen technique for creating the biological motion stimuli for the experiments reported in chapters 5 and 7. The basic technique originates from Johansson's (1973) patch-light technique. Given the available resources at the time, this method was more economically feasible, although quite labor intensive.

4.4.1 Recording Patch-light Actions

One human actor (male) was dressed in tight fitting dark clothing, and a band of reflective material (width = 20 mm) was wrapped around the major joints of his body and around the upper portion of his head. This resulted in 12 patches of reflective material. (The total here is 12 because instead of 2 markers for the hips, one band was wrapped around the hip area.) The material was wrapped around the joints in order to avoid occlusion artifacts mentioned previously. I should mention that the use of only one actor could potentially lead to biases in the displays. At the time these stimuli were created, however, manual digital editing of many action sequences performed by two actors was far too time consuming. It should also be pointed out that the purpose of the experiments was to investigate the categorization of actions. In this sense, including different actors would only likely lead to variations in the local trajectories of the patches, not the more global motion patterns. There are some results that suggest that small variations in local motion trajectories have little impact on action identification and detection. For example, Dittrich (1993) included a condition where the reflective patches were placed between the joints on an actor. The results showed that reliable identification of actions with this inter-joint condition did not differ from the standard condition where the patches were place directly on the joints of the actor. Visual identification of actions is apparently robust to variations of patch placement on an actor. Giese and Poggio (2003) also assert that an important aspect of action recognition is the ability to generalize across the identity of the actor. They report (Giese & Poggio, 2002) that simulations based on their model show clear action recognition generalization over different actors performing the same action. The gist here is that the evidence suggests that even if one actor is used to create the patch-light displays in the experiments reported in this book there is little risk that action identification or recognition will be significantly influenced by the minor idiosyncratic movements of that person.

The initial recording of actions for the experiments presented in chapter 5 were based on the objectives of exploring the graded structure of action categories as well as investigating the extent to which the categorization decisions (within and between) of subjects are affected by the perceptual similarity of action exemplars. With this in mind, I recorded a number of action exemplars from different action categories. In chapter 5, I will discuss the choice of the categories and exemplars. A selected list of recorded actions is presented in Table 4.1.

The actions were recorded with a standard video camera (Pal) with a light source attached to the camera to illuminate the actor. One of the first issues to contend with when using the direct video recording technique is the angle at which to film the action sequences. It is not likely the case that the same in-depth viewing angle is the most advantageous for all action categories or even exemplars within the same category. A further difficulty is the fact that an action sequence can involve the rotation of the body. For example, dancing, climbing a rope and throwing involve some body rotation. The guiding principle adopted when recording the actions was to maintain an acceptable trade-off between the number of visible patches and the perceived speed of the patches. In other words, the perceived speed of the patches around the wrists of an actor is influence by the viewing angle. For example, the perceived speed of the wrist patches would be faster when viewing a throwing action from a sagittal view than from a frontal view. Attempts were made to set the recording angle so that both visibility of patches and their perceived speed remained high. As far as determining the most advantageous viewpoint for viewing point-light displays, Bradshaw et al (1999, as cited by Thornton, 2005) obtained results showing the best performance for a ¾ view. A similar recording angle was used for most of the actions listed in Table 4.1.

Table 4.1. Recorded patch-light action exemplars according to action category.

Category	Exemplars
Catching	• over the shoulder • to the side (waist high) • over the head (near head level) • straight on (abdomen level)
Climbing	• onto a table • up a ladder • up a rope • up stairs (side view) • up stairs (front view)
Crawling	• on hands and knees • “army” crawl • “seal” crawl
Jumping	• straight up in the air • standing long jump • stride jump • hopping • jumping rope • onto a table • down from a table • jumping jacks
Kicking	• punt • soccer style • toe kick • heel kick • side kick (karate) • front kick (karate)
Pulling & pushing	• pushing a cart • shoving a cart • pulling a cart
Running	• sprinting • jogging • sideways • backwards • in place • skipping
Walking	• normal • with a limp • fast walk • backwards • sideways • “crab” walk • hands even w/ ipsilateral legs • high-step walk
Throwing	• overhand with a football • overhand with a baseball • sidearm • lateral toss • underhand • soccer throw-in • shooting a basket • lob • drop
Waving	• simple hand wave • whole arm wave • both arms (over head) • “come here” wave • “get back” wave

Exceptions to the ¾ view include jumping rope and performing jumping jacks, where the extent of the up-and-down motion of the arms was best captured with a full frontal view. It should, however, be pointed out that canonical viewpoints for different actions should be further investigated.

A further methodological difficulty concerns the time frame for different action sequences. Some actions are cyclical, e.g., walking and running. Others are more non-cyclical, e.g., throwing and kicking. In an attempt to standardize the duration of the sequences, I selected a reference frame that constituted the "middle" of an action. For walking, the "middle" part of the walking cycle was deemed as the point at which one leg of the walker produced the most occlusion as it past in front of the other leg. For throwing and kicking, the "middle" was deemed as the release point or point of contact with the ball. Creating the digitized sequences then proceeded from this middle point out to the remaining frames toward the beginning and end of the action sequences. The issue of how to create temporally standardized action sequences is a difficult issue. It may be the case that some actions require more time to execute. In my choice of action sequences, I tried to select actions that could be performed and recognized fairly quickly.

4.4.2 Digitizing and Editing

Once the sequences were recorded, they were then digitized frame-by-frame on a Macintosh ci. All subsequent editing was also carried out on the same computer. The digitizing technology at the time did not allow for a more automated digitizing of the sequences.[29] Since this process was labor intensive, I digitized about 3 seconds from each recorded sequence, which resulted in approximately 75 digitized still images for each action (based on the Pal standard of 25 interlaced images per second). I recorded a total of 60 actions, which resulted in roughly 4,500 digitized images for all action sequences.

The process of isolating the reflective patches from the rest of the image was carried out by creating an editing routine in Photoshop®. The basic stages of the editing routine are presented in Figure 4.1. The first step involved increasing the contrast settings and decreasing the brightness in the images. Images were then converted to black and white. The human figure was selected and cut out from the rest of the image. The reflective patches were then selected using the pixel selecting tool and then the remainder of the image was deleted. Finally, the white patches were converted to black against a white background.

---

[29] The digitizer used at the time was able to digitize "on-the-fly," but it conveniently skipped some frames in order to keep up with the real-time video sequence. There were also some problems with the digitizer's ability to lock on to the image signal, which resulted in some lost frames for the action sequences. Consequently, I reduced the captured frame rate to 20 images per second.

4.4.3 Reanimation

Because each film sequence was digitized and edited as a series of still images, the images had to be reanimated. This was done by importing the images into, MacroMind Director ®, a program used to create animated displays. At this point, the animations contained a lot of jitter due to the placement of the displays on the screen. Motion jitter was smoothed by correcting the placement of the images in relation to previous images. The original recording was also used as reference by which to compare the reanimated sequence. The animation software could also be used to manipulate the displays in various ways. For example, the displays could be rotated and resized. It was also possible to create dynamic masks based on the trajectories of the individual patches in the patch-light figure (Cutting et al., 1988).

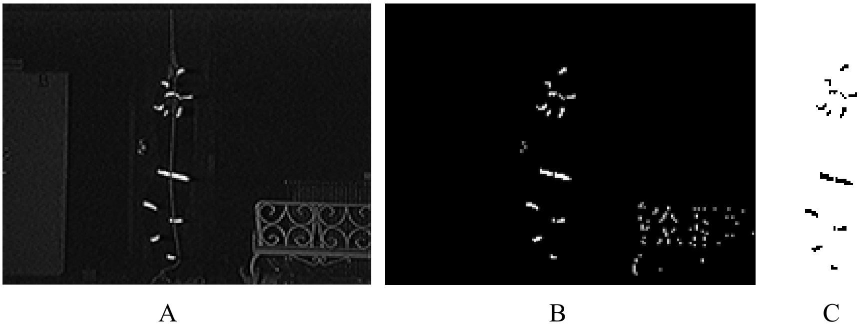


Figure 4.1. The three panels show the same image taken from a sequence of a person climbing up a rope. Panel A shows the digitized image. Some structural features are readily identifiable. Panel B is the result of changing the contrast and brightness settings. Panel C shows the figure cut out from the rest of the image and converted to black on white. The figures in the panels are all the same size.

4.4.4 Software for Presenting the Displays and Running Experiments

Given the lack of software for running dynamic stimuli in experimental settings at the time, we decided to produce our own. Christian Balkenius created DotPlayer to display the patch-light sequences as well as collect data from experiments. The basic functions of the program allowed a user to import a series of images (PICS) from a file database and to position the images on the computer screen. An experiment could be fairly easily set up using standard trial features, e.g., inter-stimulus intervals, fixation points, primes, etc. One important function of the program, however, concerned manipulation of the display rate and the inclusion of every image included in the image sequence. Unlike other programs designed to run dynamic displays at the time, this program maintained the integrity of the displays by making sure it showed each full frame. As a program for running experiments, DotPlayer was also able to randomize stimulus presentation and log responses, including reaction time, from a standard Macintosh keyboard. Of course there are now numerous powerful software packages that allow a user to create many different kinds of experiments that use a variety of different stimuli and stimuli formats.

4.5 Future Developments in Techniques for Studying Action Recognition

Given the recent advances in hardware and software as well as increasing research interest, the creation of point-light displays has become more accessible. There are also currently publicly available point-light databases where researchers can download stimuli for use in experiments or demonstrations. Ma, Paterson and Pollick (2006) have developed a library of 4080 human movements based on motion capture data. An important purpose of the library is to provide stimuli that can be used to study the sources of variability in human movements. It contains actions from 30 individuals performing kicking, throwing, knocking and lifting actions. It also contains actions that express affective styles (sad, angry, happy, and neutral). The access to motion capture data also allows action to be viewed from different viewpoints. Another feature of the library is that solid body as well as point-light displays can be created from the motion capture data. The data require, however, specific software (Character Studio) to view the displays.

Vanrie and Verfaillie (2004) created an accessible library of 22 actions using motion capture. Each action is viewable from 5 different viewpoints, which allows for the systematic investigation of viewpoint dependence. In contrast to the Ma et al. (2006) library, the actions in the Vanrie and Verfaillie library were recorded using the same actor. One nice feature of this library is the format. All actions are available in .avi-format, which makes them relative easy to display. It should be noted, however, that there is no natural occlusion (no explicit depth cues) in these files.

Shipley and Brumberg (2005) used a markerless motion capture technique based on extracting the 2D coordinates of the joints of human and some nonhuman movements. Over 90 point-light displays are included in the library. While the library includes downloads of low image quality Quicktime movies, it also includes data files of the 2D coordinates for each action as well as software for running the files. In contrast to the actions in the library of Vanrie and Verfaillie (2004), the actions created by Shipley are not available from systematically different viewpoints.

Point-light techniques represent an effective method for isolating motion cues in human actions. As Thornton (2006) mentions, however, developing a better understanding of how we perceive the actions of others will most likely involve investigating not only motion cues but the interaction between motion and form cues. In this regard, it is important to note that Thornton (2006) discusses the development of techniques for looking more closely at the interaction between form and motion cues using actions recorded in a natural environment. An example of such a technique is used in Vuong, Hof, Bülthoff and Thornton (2006) where they superimposed video recorded target sequences on distracter sequences. A central question in their research concerned the ability of subjects to detect a person walking (target) in static and dynamic modes of presentation. The results showed a clear benefit for target detection if the walker was presented as a dynamic target. Future work in this area needs to more precisely address the nature of dynamic information and the role of form cues in segmenting objects from one another in natural scenes.

Chapter 5 - Action Categories: Graded Structure, Prototypes and Context Effects[30]

For object categories, it appears that a radial structure around a salient prototype reflects an important aspect of a category's psychological organization. Converging evidence for (proto)typicality effects within various categorical domains has been obtained with a number of different measures, e.g., instance dominance, category dominance, goodness of exemplar ratings, word/item frequency, familiarity and reaction time (RT) in the context of speeded verification tasks (Komatsu, 1992). Perhaps the most salient and robust relationship between these different measures is the typicality-RT effect. When subjects are given the task of rating the representativeness of various category exemplars (different kinds of vehicles) in relation to a higher-level category label (VEHICLE), this variable serves as a reliable predictor for the time it takes subjects to correctly verify CATEGORY–[exemplar] relations, e.g., VEHICLE-[car]. Subjects verify the category relations containing the more representative exemplars faster than relations containing less representative exemplars.[31] Casey (1992), for example, found that typicality, as measured by goodness-of-exemplar ratings, was the best predictor of verification-RT for *yes* responses among the other variables (instance dominance, word frequency and category dominance) in the experiments reported there. The gist of these findings is that in addition to the prevalence of the typicality-RT effect within different experimental paradigms (Rosch & Mervis, 1981), it is also unique with regard to other variables that are correlated with verification response time.

In addition to the finding that verification RT varies as a function of typicality for the *yes* (true) responses, a context effect has been found to occur among the *no* (false)

---

[30] Portions of this chapter appear in Hemeren (1997).

[31] See, however, Murphy and Brownell (1995) for exceptions to this finding when highly atypical exemplars are used.

responses (McCloskey & Glucksberg, 1979). In the course of correctly responding to true items, e.g., 'A robin is a bird,' subjects are also given the task of correctly responding to false item pairs, which can vary in their *perceptual relatedness* according to form. An example of an *unrelated false* item pair would be 'A chair is a bird,' in contrast to a *form related* false item pair, e.g., 'A bat is a bird.' The context effect occurs when subjects respond significantly faster to the *unrelated* false items than to the *related* false items. Like the typicality-RT effect, the context effect has been well replicated, for example in the work of Casey and Heath (1989).

In chapter 2, I presented previous findings suggesting that action categories are similarly organized according to a radial structure around a central representation, i.e., prototype. If this is the case, then we may see similar effects of typicality and context (perceptual similarity) on category verification times for actions. In relation to the general thesis of this book, the purpose of the experiments in this chapter is to further investigate the categorical knowledge associated with human action perception. More specifically, the first experiment in this chapter attempts to further our understanding of the psychological organization of action categories by asking subjects to rate the typicality of action exemplars in relation to a previously presented category label. This will be done for correct category-exemplar sequences, e.g., WALKING (target category)→[marching] as well as for incorrect category-exemplar pairings where the exemplar comes from a contrast category, e.g., THROWING (contrast category)→[marching]. Evidence for the graded structure of action categories will be obtained if subjects provide different typicality ratings for action exemplars from the same category, e.g., WALKING→[marching] vs. WALKING→[limping]. In addition, the first experiment will investigate the extent to which subjects view actions as being typical of contrast categories. The extent to which an action exemplar is viewed as being at least somewhat typical of a contrast category depends on the perceptual similarity between an action exemplar and the representation (prototype) for a contrast category. Another likely factor is the extent to which a viewer is sure about what he/she sees. If, for example, there is some doubt about the identification of an action exemplar, subjects may view the action as being somewhat typical of a contrast category that is perceptually similar to the target category, e.g., WAVING and THROWING. If, on the other hand, an action exemplar is distinct in the sense that there is little doubt about the identification of the action, it should have less of a chance as being rated as somewhat typical of a contrast category, even when the contrast category may be perceptually similar to the target category.

The purpose of the second experiment is to investigate the extent to which the possible radial structure of the action categories is revealed by a speeded category verification task. According to the typicality-RT effect found for object categories, the time it takes to verify the category membership of an action exemplar is correlated with the judged typicality of the exemplar. If, when presented with a category label, a prototype representation is accessed for the category, then category verification of a following action exemplar of that category will depend on the similarity between the prototype and the action exemplar. Action exemplars that are more similar to the accessed prototype, i.e., more typical, will be verified faster than more atypical action exemplars.

An additional issue in regard to the category verification task is the extent to which the perceptual similarity of instances from *contrast categories* will influence the time it takes to verify that an action exemplar is/is not a member of a certain category. Similar to the reasoning above for typicality judgments, the more perceptually similar an action exemplar is to members of a contrast category the longer it should take to verify that it is not a member of the contrast category.

In contrast to the task of judging typicality, the category verification task assesses judgments of category membership. It may very well be the case that people can see actions from contrast categories as being somewhat typical of the contrast category (a kicking instance does not belong to the category of running, but it may nonetheless be viewed as an atypical instance of running, e.g., dribbling a soccer ball). The point here is that the fact that an atypical action might be seen as somewhat typical of a contrast category does not necessarily mean that the subject views it as a *member or instance* of the contrast category. This issue is important to keep in mind because while perceptual similarity is an important factor in determining the typicality and membership of category exemplars, there may be other aspects of conceptual knowledge that also play a role in judgments of typicality and category membership.

5.1 Experiment 1: Typicality Ratings[32]

Although typicality effects have been found for a wide range of stimulus material (Barsalou, 1987), it appears that the context effect has been largely restricted to natural kind and artifact categories based on instance dominance norms, as for example, collected by Battig and Montague (1969). The purpose of this experiment is to gather typicality (goodness-of-exemplar) ratings for a number of different actions in relation to a previously presented category label. The strategy here, in contrast to presenting words that denote category *exemplars*, is to present subjects with patch-light displays depicting various ways of KICKING, RUNNING, THROWING and WAVING. Even if subjects do not know what the different ways are called, they should still be able to recognize a soccer-throw-in as an instance of THROWING. If they succeed in this task, they should also be able to provide a rating as to how typical it is of the category THROWING. Two issues will be investigated in this experiment:

---

[32] It should be noted that the presentation of the experiments does not match the chronological order in which the experiments were carried out. Experiment 2 in this chapter was actually carried out before experiment 1. The reason for presenting the experiments in the present order is that this order reflects a better conceptual ordering of the experiments in relation to the theoretical issues mentioned in chapter 2.

1) the extent to which judgments of typicality reveal graded structure for 4 action categories and 2) the extent to which action exemplars of relatively low typicality might also might be viewed as being at least somewhat typical of contrast categories that are perceptually similar to the action exemplars. The obtained typicality rating will then be compared with the verification-RT data in the following experiment in order to assess the relation between the two variables in the context of a typicality-RT effect.

5.1.1 Method

5.1.1.1 Subjects

Twenty-four native Swedish-speaking students (11 females, 13 males) from Lund University participated in the experiment (mean age = 24). All subjects had normal or corrected-to-normal vision and received no compensation for participating in the experiment.

5.1.1.2 Materials and Design

Twenty actions were filmed as patch-light displays. The actions correspond to 5 different ways of KICKING, RUNNING, THROWING, and WAVING. The list of selected action exemplars is presented in Table 5.1. Since there is previously little specific research on the cognitive organization of action categories, the choice of action categories and action exemplars within each category was based partly on the results from response frequencies and mean ordinal positions presented in Chapter 3. As a first step in conducting empirical studies in action categorization, I chose actions that were listed quite often and occurred relatively early in the subjects' lists. A further criterion was to choose action categories that had a clear perceptual salience either in whole body movement or in the distinct movement of a body limb. It was necessary to choose action categories where different ways of performing the actions (subordinate action exemplars) would be at least intuitively somewhat familiar and could be filmed as patch-light displays. This allowed the creation of actions exemplars that reflected different manners of performing an action. A further aspect that influenced the choice of action category was the issue of relatedness. While RUNNING and KICKING are perceptually related via leg motion, THROWING and WAVING are perceptually related via arm motion. Perceptual relatedness can give rise to a context effect when subjects are asked to verify the categorical relation between a previously presented category label and an action exemplar that does *not* belong to the category but is perceptually similar by use of body part to a prototype (or template) activated by first being presented as a category label. This effect will be investigated in exp. 2.

Finally, the different ways of performing the actions were selected on the basis of an 'intuitive' understanding of how they might differ with regard to representativeness for each one of the categories but yet be classified as belonging to that category. Since the primary purpose of this first experiment is to investigate the graded structure of four different action categories, I selected five actions that I thought might differ in regard to their typicality. The one possible counter example to this constraint is skipping. I was unsure whether or not subjects would regard it as running or not. Skipping could be viewed as a kind of hop or jump as well as a kind of running motion.

Table 5.1 Action categories and action exemplars used in experiments 1 and 2.

	Action categories			
	Kicking	Running	Throwing	Waving
Action exemplars	heel-kick	sprint	throw-in	arm
	karate-kick	backwards	side-toss	both arms
	punt	in place	sidearm	come here
	soccer-kick	sideways	overhand	get back
	toe-kick	skip	underhand	hand

The patch-light displays were created using the method described in Chapter 4. The displays were edited to remove translation motion. Hence, the actions were presented as if the camera were following the actions across a surface. Each of the 20 action sequences consisted of 15 frames. The starting and ending points for the kicking and throwing actions were determined by locating frames for the point of release for the throwing actions and the point of contact for the 3 kicking actions that involved kicking a ball. For these actions, the point of release and point of contact frames served as the middle point of each sequence. The total of 15 frames was then obtained by selecting 7 frames prior to the middle point and 7 frames following the middle point. For the karate kicks, the middle points were defined as the point at which the kicking leg makes a maximal extension for that kind of kick. Seven frames prior to and following the middle points were then included to create each 15-frame sequence. Frame selection for the different waving actions was determined by finding the 15 frames that included the largest portion of waving motion. For the running actions, the 15-frame sequences were selected in order to include as much of a complete cycle as possible. In this case, I defined a cycle as the interval in which the body moves between two similar support phases. In the case of the sprinting action, however, the sequence is approximately 2 frames short of a complete cycle. Figure 5.1 contains the starting, middle and ending frames for one action from each of the action categories used in this experiment. Note that the lines connecting the patches were not visible during the experiment. The lines are presented here in order to better illustrate the human figure in the static frames.

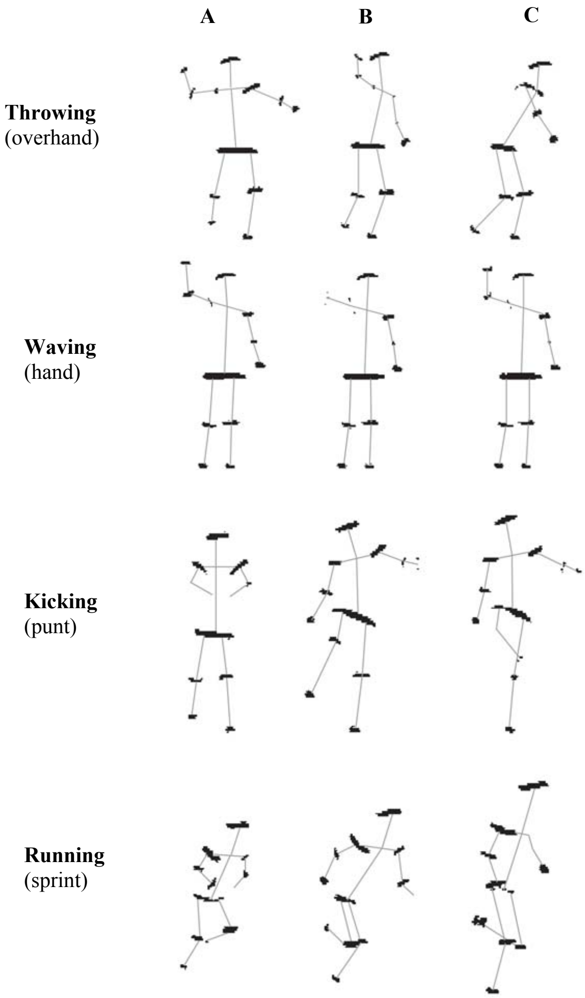


Figure 5.1. Examples of the patch-light stimuli used in experiment 1. The letters A, B and C refer to the beginning, middle and end frames respectively. The lines connecting the patches are for illustrative purposes only and were not included in the experiment.

The stimulus displays were presented on an Apple 6100/66 power PC. DotPlayer was used to present the stimuli and record subject responses. The patch-light sequences were centered in the display and presented at a rate of 20 frames-per-second and subtended a visual angle of approximately 6 to 7° in height. The duration of each patch-light action sequence was 750ms.

The experiment includes 3 independent variables; Category-exemplar agreement (matching and nonmatching), Action category (KICKING, RUNNING, THROWING, and WAVING) and Action exemplar (5 different exemplars per action category). The Exemplar variable is nested within Action category for the matching level of the Category-exemplar agreement variable. For the nonmatching level, however, action exemplars are crossed with the different action category labels to create all possible combinations of action category labels and action exemplars. This creates a nonsymmetrical design where there are only 20 possible combinations of matching, action category and exemplar levels (see Table 5.1), whereas there are 60 possible combinations of nonmatching, action category, and exemplar levels.

5.1.1.3 Procedure

Three blocks of 20 *nonmatching* trials were constructed to include all possible *nonmatching* combinations of the category labels and actions (Table 5.2).

Every subject participated in 3 consecutive sessions separated by a short break. Each session consisted of viewing 20 *matching*-trials and one block of 20 *non–matching*-trials. Over the course of the 3 sessions, each subject viewed each block of nonmatching trials once and viewed the matching trials 3 times each. Since there were 6 possible orders in which subjects could view the nonmatching trials, a block randomization schedule was prepared for all orders of presentation. Subjects were randomly assigned to the different block orders such that each nonmatching block of occurred an equal number of times (4) for the experiment as a whole. Trials within each session were presented in a random order.

Table 5.2. Construction of blocks for nonmatching trials.

Blocks of nonmatching trials		
Block 1	Block 2	Block 3
THROWING → kicking	THROWING → running	THROWING → waving
KICKING → running	KICKING → waving	KICKING → throwing
RUNNING → waving	RUNNING → throwing	RUNNING → kicking
WAVING → throwing	WAVING → kicking	WAVING → running

The category labels and actions were presented in a category(WORD)-exemplar order, e.g., KICKING–kicking exemplar. The time line of the trials was as follows: fixation point (1000ms)–ISI (500ms)–category label (2000ms)–ISI (500ms)–action exemplar (750ms). The category labels and the action instances were both centered on the fixation point. The action exemplar could only be viewed once. There was no way to repeat a given trial.

A modified Likert scale based on Kalish (1995) was used as a reference for the typicality ratings (see Figure 5.2). By using this Likert scale, subjects were given the option of rating typicality in absolute terms. This was done to minimize a potential bias towards giving graded typicality ratings. The scale was visible to the subjects throughout the experiment. It was printed on paper and taped on the bottom of the computer screen.

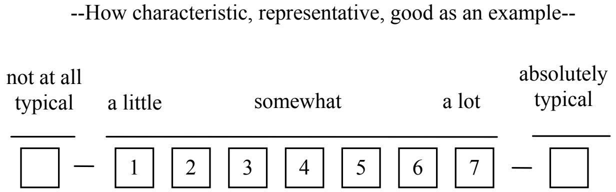


Figure 5.2. Modified Likert scale based on Kalish (1995).

Subjects were tested individually and were seated so that viewing distance was approximately 70cm to the computer screen. They were informed about the sequence of stimuli in the trials. They were also instructed to read the category label quietly to themselves and then make a judgment as to how characteristic, typical, good as an example the different actions were in relation to the preceding category label. The subjects responded by pressing the key on the keyboard that corresponded to their rating according to the scale in front of them. The 'not at all typical' and 'absolutely typical' keys were represented by colored tape on the keys directly to the left and right of the number 1 key and the number 7 key respectively, on a Swedish keyboard. Trials were self-paced in the sense that proceeding to the next trial required a response from the subject.[33]

---

[33] In addition to the task of providing typicality ratings, subjects were also instructed to provide judgments of category membership, e.g., how good of member an action exemplar was in relation to the presented category label. This was done to assess the extent to which judgments of typicality and category membership are dissociable. Since this issue is not a focus of the book, I will not present any data for judgments of category membership. In terms of procedure, judgments of category membership always followed typicality ratings.

Prior to the test trials, the subjects were acquainted with the task by having them participate in a practice session consisting of 28 trials. The stimuli used in the practice trials were not used in the test trials.

5.1.2 Results and Discussion

Since the first issue in this experiment has to do with the graded structure of action categories for instances where typicality is judged in relation to relevant category labels, results for those trials are presented first. Results for the nonmatching trials and their comparison to the results for the matching trials will follow. In short, the results are presented in two sections, i.e., matching and nonmatching trials.

5.1.2.1 Matching Trials

The 3 typicality ratings for the *matching* trials from each subject were averaged to form a mean for each subject and action exemplar. Table 5.3 presents the mean typicality ratings for the *matching* conditions for the five action exemplars in each action category. For all action categories, it appears that the different action exemplars differ to some extent in their typicality. Analyses of each action category will be presented below. One other point to notice about the general pattern of results is the fact that the typicality ratings are all rather high, i.e., all above 4, which indicates that all action exemplars were judged at least somewhat typical in relation to the relevant category labels. This result could reflect a bias (possibly a demand characteristic) to judge all actions as being somewhat typical of a given category. This explanation, however, seems unlikely given the data for nonmatching trials, which were randomly interspersed with the matching trials in each session. As will be discussed below, subjects did in fact use the full range of the Likert scale in their judgments.

Separate repeated measures analysis of variance (single factor) (ANOVAs) were carried out for each category. The main point of interest was whether or not there were significant differences among the mean typicality ratings for the action exemplars within each category and not whether the categories differed from one another. Following the detection of an overall significant difference among the means, 10 post-hoc pairwise comparisons between the different subordinate level action exemplars were carried out in order to detect which means were different from each other. The reason for these comparisons was to see if there were any 3 actions within each category that would constitute relative high, medium and low actions of typicality. The pairwise comparisons are adjusted for Type I error by a Bonferroni adjustment, which put the alpha level of significance at .005 for all such comparisons.[34]

---

[34] The reported results are based on the univariate tests unless otherwise noted.

Table 5.3. Mean typicality ratings (Typ.) and standard deviations (SD) for action exemplars in relation to an action category label.

	Action category label							
	Kicking	Typ.	Running	Typ.	Throwing	Typ.	Waving	Typ.
Action exemplar	soccer	7.67 (.6)	sprint	7.87 (.5)	overhand	7.18 (1.2)	hand	7.76 (.6)
	punt	7.60 (.8)	skip	5.01 (2.1)	throw-in	6.75 (1.1)	both arms	5.31 (1.8)
	toe-kick	7.25 (.9)	backwards	4.29 (2.1)	side arm	6.64 (1.4)	get back	5.22 (1.7)
	karate	6.56 (1.5)	sideways	4.31 (2.1)	underhand	5.94 (1.5)	come here	4.61 (2.1)
	heel-kick	4.83 (1.9)	in place	4.03 (2.2)	side toss	5.83 (1.2)	arm	4.29 (2.0)

Kicking. The results for the kicking category revealed an overall significant effect of action exemplars [*F*(4,92) = 34.99, MSE = .95, partial η^2 = .603, $p < .0001$]. The pairwise comparisons (Bonferroni adjustment, $p < .005$) show that the typicality ratings for SOCCER were significantly greater than for KARATE, and the typicality ratings for KARATE were significantly greater than for HEEL-KICK [*F*(1,23) = 17.36, MSE = 1.71, partial η^2 = .430; *F*(1,23) = 26.92, MSE = 2.64, partial η^2 = .539, respectively]. These results indicate a graded structure for the KICKING category with SOCCER, KARATE and HEEL-KICK corresponding to levels of high, medium and low typicality respectively.

Running. The effect of subordinate level action was also significant for the RUNNING category [*F*(4,92) = 41.34, MSE = 1.47, partial η^2 = .643, $p < .0001$]. The results of the pairwise comparisons indicate that SPRINTING was rated as more typical than SKIPPING, which was rated as more typical than running SIDEWAYS [*F*(1,23) = 39.64, MSE = 4.96, partial η^2 = .6330; *F*(1,23) = 12.89, MSE = .89, partial η^2 = .359 respectively]. As in the KICKING category above, there is a clear typicality gradient among the action exemplars.

Throwing. The subordinate level action effect was significant for the THROWING exemplars, [*F*(4,92) = 7.11, MSE = 1.09, partial η^2 = .236, $p < .0001$]. Unlike the previous results for KICKING and RUNNING, the pairwise comparisons show only a two-tiered typicality gradient that reached significance in the context of the Bonferroni adjustment. OVERHAND was rated as more typical than SIDE TOSS and UNDERHAND [$F(1,23) = 16.31$, MSE = 2.67, partial $\eta^2 = .415$; $F(1,23) = 13.39$, MSE = 2.74, partial $\eta^2 = .368$ respectively]. THROW-IN was also more typical than SIDE TOSS [$F(1,23) = 14.47$, MSE = 1.39, partial $\eta^2 = .386$]. No other post-hoc comparisons reached significance according to the Bonferroni adjustment.

Waving. Differences between the WAVING exemplars were significant [$F(4,92) = 28.24$, MSE = 1.59, partial $\eta^2 = .551$, $p < .0001$]. The paired comparisons indicate only that HAND is significantly more typical than the other WAVING exemplars, all *p*s < .005.

For the action categories in this study, a three-tiered difference between the rated typicality for action exemplars was found for KICKING and RUNNING, while only a two-tiered difference was found for THROWING and WAVING. Similar to object categories, these results suggest that for a restricted domain and number, typicality ratings for actions show graded structure. This, however, should not be surprising given previous findings about the robustness of typicality effects for a broad range of categorical domains (Murphy, 2002). Indeed, a finding showing no graded structure for action categories would have been more remarkable.

5.1.2.2 Nonmatching Trials

The *nonmatching* items were constructed by presenting all action exemplars in the context of the *nonmatching* (contrast) category labels, e.g., THROWING–running exemplars. The mean typicality ratings for the nonmatching trials are presented in Table 5.4.[35] In all but 5 of the 60 conditions, mean typicality ratings were less than 1. Subjects frequently used the option of “not at all typical” on the modified Likert scale. Mean typicality ratings greater than 1 were obtained for the following action exemplars: *arm* (4.58), *come here* (3.33) and *get back* (1.29) in the context of the THROWING category label, and *sidearm* (1.50) and *overhand* (1.12) under the WAVING category label.

It appears that subjects were categorical in their typicality ratings of action exemplars when initially presented with a contrast category label. Relative to the other action exemplars within an action category, some actions received a low typicality

---

[35] All responses of “not at all typical” were given a value of zero. The basic idea is that if an action exemplar is “not at all typical” it has zero typicality in relation to the contrast category. Given the extent of subject agreement for using the “not at all typical” alternative, there is virtually no variance associated with the means, which rules out evaluating differences between the means using analysis of variance and makes the reporting of standard deviations uninformative. This, however, does not mean that the data are uninterpretable. The fact that there was very little variance shows that subject agreement regarding the typicality of action exemplars in relation to contrast category labels was very high.

rating during the matching trials (above). Despite this difference, subjects were *not* inclined to rate the action as being typical of a contrasting action category. The exceptions to this trend were restricted to action exemplars that were perceptually related, e.g. waving and throwing. Kicking and running actions were also perceptually related in relation to leg movement, but did not exceed a mean typicality rating of 1 when presented in the context of a related contrast category label.

Despite this result, all mean typicality ratings for running and kicking exemplars when presented in the context of a related contrast category label (KICKING or RUNNING) were greater than 0, indicating that category relatedness may contribute to increased typicality for exemplars from contrast categories. This result appears to be noticeably weaker when running and kicking exemplars were preceded by category labels for non-related categories, e.g., THROWING and WAVING. This issue of relatedness will be specifically addressed in experiment 2.

Although not specifically tested in this experiment, the results can be related to the issue of basic level actions mentioned in chapter 3. The salient difference in typicality ratings for matching and nonmatching trials lends some support to previous findings for object categories where objects categorized at the basic level are maximally more distinct from one another than objects categorized at the subordinate level. Given the few exceptions in the nonmatching trials, subjects viewed most of the actions as “not at all typical” of the contrast category labels. This suggests that the chosen action categories reside on a basic level in an action category hierarchy.

The last point to address before moving on to the next experiment is to offer an explanation of why 5 actions received typicality ratings greater than 1 in the nonmatching trials. Recall the results from Pollick et al. (2001) showing that categorization judgments improved as the exaggerations moved further away from the grand average. For the dissimilarity judgments in their experiment, the results showed that increasing the spatial exaggerations in the movements resulted in a corresponding difference in the distance between the movements in psychological space. The point here is that the confusability of action exemplars with a contrast category is both a function of judged typicality and distinctiveness. While an action (arm wave) is given a low typicality rating (4.29) relative to the other waving exemplars, it may not be as distinctive in terms of its spatiotemporal pattern as other waving exemplars (get back). This may explain why the arm wave was viewed as being more typical of throwing than the get back wave (or even both arms) despite no significant difference in typicality between the actions in relation to the waving label.

Table 5.4. Mean typicality ratings for action exemplars in relation to a contrasting action category label. Note: Boldface type is used to highlight means greater than 1. Mean typicality ratings greater than 1 are highlighted in bold-faced text.

	Contrasting category label														
							Throwing								
Exemplar	heel-kick	karate	punt	soccer	toe-kick	sprint	backwards	in place	sideways	skip	arm	both arms	come here	get back	hand
	.17	.37	.08	.08	.33	.29	.00	.00	.25	.25	**4.58**	.08	**3.38**	**1.29**	.29
							Kicking								
Exemplar	throw-in	side toss	side arm	overhand	underhand	sprint	backwards	in place	sideways	skip	arm	both arms	come here	get back	hand
	.00	.00	.42	.04	.08	.29	.25	.54	.29	.79	.00	.04	.00	.00	.00
							Running								
Exemplar	heel-kick	karate	punt	soccer	toe-kick	throw-in	side toss	side arm	overhand	underhand	arm	both arms	come here	get back	hand
	.17	.21	.54	.37	.29	.08	.00	.04	.04	.00	.00	.00	.00	.00	.00
							Waving								
Exemplar	heel-kick	karate	punt	soccer	toe-kick	sprint	backwards	in place	sideways	skip	throw-in	side toss	side arm	overhand	underhand
	.08	.04	.00	.08	.04	.04	.04	.04	.04	.04	.92	.62	**1.50**	**1.12**	.79

5.2 Experiment 2: Category Verification

The purpose of this experiment was to investigate the structure of action categories with regard to the time it takes subjects to verify whether or not an exemplar belongs to a previously presented category according to a category label. If the typicality ratings in the previous experiment and the verification reaction times in this experiment are a function of the same process, i.e., judging the similarity between an exemplar and a category prototype, then they should be highly correlated.

In addition to this typicality-RT effect, actions that are perceptually similar but categorically different should lead to longer verification times when presented in the context of a perceptually related but categorically different category label. As mentioned at the beginning of this chapter, this effect is well established for object categories (McCloskey & Glucksberg, 1979; Casey & Heath, 1989).

For the action categories used in these two experiments, KICKING and RUNNING are perceptually similar in regard to the locus of motion centering on the legs, and THROWING and WAVING are similar in regard to the motion of the arms. The results from experiment 1 showed that 5 action exemplars were judged as being at least somewhat typical of a contrast category. This finding was restricted to 3 waving exemplars and 2 throwing exemplars. It was also the case that running and kicking exemplars were viewed by some subjects as being “a little” typical of perceptually related contrast categories, i.e., KICKING and RUNNING respectively. Using a speeded verification task in conjunction with repeated measures of the same condition, the current experiment may be more sensitive to the effects of relatedness than the previous experiment where judgments of typicality were more coarsely measured. If this is the case, then we should see context effects for these categories as well.

If category verification of an action exemplar is carried out by accessing an action prototype for the action category label, then highly typical actions should be verified faster than less typical actions. This should result in the typicality-RT effect. Along similar lines, less typical actions may also share more properties (spatiotemporal pattern) with a prototype from a contrast category, which will result in longer verification times and/or more verification “errors.”

The action categories can also be viewed from the perspective of functional similarity. Although perceptually different, KICKING and THROWING, for example, are functionally related in the context of propelling an object (Schank & Abelson, 1977). Given the nature of the stimuli, one would expect the perceptual similarity of the different motions to affect verification-RT for related categories along this dimension. The notion that these action categories may be related in terms of the goal of performing a certain movement is less apparent. This experiment is also intended to address this issue of functional relatedness between action categories.

5.2.1 Method

5.2.1.1 Subjects

Twenty-one native Swedish-speaking students (12 male, 9 female) from Lund University participated in the experiment (mean age = 24). All subjects had normal or corrected-to-normal vision and received no compensation for participating in the experiment. No subjects in this experiment participated in the previous experiment.

5.2.1.2 Materials and Design

The patch-light displays and trial construction, including stimulus duration, used in the previous experiment were also used in this experiment. DotPlayer was used to present and record response times to the nearest millisecond. The trials were presented using an Apple IIci.

Three blocks of trials were constructed for this experiment. The primary purpose of the three blocks was to create a series of trials that could be repeated a number of times (5) in order to obtain stable reaction times, i.e., reduced error variance, for the verification task. Recall, as in experiment 1, that in contrast to the 20 matching trials there were 60 possible parings of nonmatching trials. For the experiment reported here, these 60 pairings were divided into 3 separate groups such that each block consisted of 20 *yes*-trials and 20 *no*-trials, i.e. parings of action exemplars with contrast categories. The 20 *yes*-trials were the same for each block and consisted of trials where the action exemplar matched the previously presented category label, as in experiment 1. The *no*-trials however, were different for each block. The *no*-trials were also further divided into 10 *related* versus 10 *unrelated* conditions. Table 5.5 contains the blocking pattern for the *no*-trials. For blocks 2 and 3, relatedness was defined by the *perceptual similarity* between an action exemplar and a prototypical action that might be accessed by a contrasting category label. For example, the waving exemplars may be perceptually related to a prototype for the THROWING category in terms of body part motion, i.e., arm movement. Recall that there was some evidence of this in experiment 1. Similarly, kicking exemplars might be perceptually similar to a prototype for the RUNNING category in terms of the movement of the legs. In contrast to the *related no*-trials, kicking exemplars are perceptually *unrelated* to a prototype for the WAVING category. If this is in fact the case, then according to the context effect, it should take subjects longer to verify *related* category-exemplar conditions than *unrelated* category-exemplar conditions.

While *perceptual* similarity determined the relatedness conditions in blocks 2 and 3, relatedness for block 1 was determined on the basis of *functional* relatedness. Throwing and kicking may involve different body parts but may be functionally similar in the sense of propelling an object. If this information is used in category verification, then it may take subjects longer to verify that kicking exemplars do not belong to the category of THROWING than for them to verify that running instances do not belong to the category of WAVING.

Table 5.5. Relatedness conditions (*no*-trials) for Experiment 2. Category labels are in capital letters. The arrow indicates the direction of verification. Category labels were presented first, followed by either related or unrelated false exemplar actions.

Block	Related Conditions	Unrelated Conditions
1	THROWING → kicking	WAVING→ running
	KICKING→ throwing	RUNNING→ waving
2	RUNNING→ kicking	THROWING→ running
	WAVING→ throwing	KICKING→ waving
3	KICKING→ running	WAVING→ kicking
	THROWING→ waving	RUNNING→ throwing

The resulting design for administering the *no*-trials is a 3 (block, between groups) x 4 (action category label/prime, within) x 5 (action exemplars nested within the category label/primes) design, resulting in total of 60 conditions.

The design for administering the *yes*-trials is much simpler due to the fact that there are no conditions for contrasting the category verification of action exemplars with other categories than the “correct” ones. In this case the design is a 4 (action category, within groups) x 5 (action exemplars nested within each category) design.

5.2.1.3 Procedure

Each participant was randomly assigned to one of the three blocks of trials. Upon arrival to the experiment room, subjects were given some information about the general nature of the task and that they would be viewing patch-light displays of biological motion. Subjects were then told they would first see a fixation point in the center of the computer screen and then a word that denoted an action category would appear. Subjects were instructed to silently read the word. Following the presentation of the word, an action exemplar in the form of a patch-light display would appear, and subjects were instructed to verify whether or not they thought the action exemplar belonged to the previously presented action category. The subjects were asked to respond as quickly as possible while keeping errors to a minimum. They were also shown the appropriate response keys on the keyboard.

The subjects responded by pushing either the right or left arrow key on the key board with the appropriate index finger. Since the “yes” and “no” responses were balanced for handedness, half of the right-handed subjects (n=8) used their right index finger for the “yes” response, and the other half (n=8) used their left index finger. Key presses were also semi-balanced for the left-handed subjects (n=5). Three left-handed subjects used their left index finger for “yes” response, and two used their right index finger.

Reaction time was measured from the onset of the action exemplar portrayed in the patch-light display to when a key press was made. The correctness of the key presses was recorded manually by the experimenter.

Subjects were seated approximately 65-70cm from the computer screen, and this distance was maintained throughout the experiment. Each participant participated in 5 consecutive sessions consisting of the same block of trials. There was a short break between sessions. Trials within the blocks and session were presented randomly.

A series of practice trials preceded the test trials to familiarize subjects with the nature of the task. No stimuli used in the practice trials were included in the test trials. The subjects were not given any feedback.

5.2.2 Results and Discussion

Following the advice of Shoben (1982), the untransformed reaction times were used in the analyses presented here. The correct reaction times for each subject and condition across the 5 sessions were averaged to form a mean RT for that subject and condition. While the cell means for most of the subjects and conditions were based on at least 3 correct reaction times, 12 cells contained 2 or fewer correct reaction times. Four cells were empty and were replaced with the mean for the condition across subjects. The cell means for conditions containing only 1 or 2 reaction times were calculated using those values. Similar to the analyses of the results in the previous experiment, the analyses here are divided into separate analyses for the *yes*- and *no*-trials. The mean reaction times for the *yes*-trials as a function of category label and action exemplar are presented in Table 5.6.

Prior to presenting the quantitative analyses, a qualitative perusal of the verification means for the different action exemplars within each action category indicates that they are different, although less so for the throwing exemplars. Subjects seem to respond faster and with fewer errors for some action exemplars compared to other action exemplars within each category.

A point worth noting is the sizable error rate for some of the action exemplars. It appears that some action exemplars were viewed as borderline cases in relation to the presented category label. In these cases, subjects may not be committing errors in a strict sense, but rather indicating that an action exemplar did not belong to the action category. An indication of this is the obtained correlation coefficient for the relation between speed and accuracy, $r = .75$ [$F(1,18) = 22.95$, $p < .001$]. This positive and high coefficient indicates that longer verification times are likely due to the relative difficulty of making a categorization decision for the action exemplars.

Table 5.6. Mean verification RTs (in msec), standard deviation (SD) and error rates (percent) for the *yes*-trials in Experiment 2 as a function of category label and action exemplar. Action exemplars are listed in the same order as the typicality ratings in Table 5.3 for easier comparison.

		Action category label			
		Kicking	Running	Throwing	Waving
Action exemplar		soccer	sprint	overhand	hand
	RT	781	678	835	800
	SD	(234)	(200)	(323)	(244)
	Error %	3	1	5	3
		punt	skip	throw-in	both arms
	RT	869	912	875	884
	SD	(260)	(517)	(225)	(297)
	Error %	10	7	5	7
		toe-kick	backwards	side arm	get back
	RT	803	936	850	1005
	SD	(255)	(291)	(263)	(414)
	Error %	6	14	4	7
		karate	sideways	underhand	come here
	RT	838	932	900	869
	SD	(250)	(347)	(266)	(296)
	Error %	10	9	5	9
		heel-kick	in place	side toss	arm
	RT	1003	1023	921	938
	SD	(317)	(381)	(370)	(616)
	Error %	20	26	9	8

5.2.2.1 Typicality-RT Effect

The pattern of results for the verification times indicates graded structure for the action categories used in this experiment. As a test of the relationship between the obtained typicality ratings from the previous experiment and the verification times in this experiment, these two measures were used in a correlation analysis. The results for this typicality-RT effect are presented in Figure 5.3.

The results show a significant correlation between rated typicality and verification reaction time, $r = -.82$, [$F(1,18) = 35.64$, $p < .0001$]. Indeed, typicality seems to be an excellent predictor of the time it takes to verify category membership for the actions used in this study. The more typical an action exemplar is rated in relation to a "correct" category label, the less time it takes to correctly verify category membership. In the next section, the *yes*-trials for each exemplar are evaluated with regard to the existence of a typicality gradient within each action category.

5.2.2.2 Yes-trials

The purpose of the analysis of the yes-trials is to assess the extent to which category verification times differ for the different action exemplars within each action category. The typicality-RT effect suggests that these two variables are related. The purpose here, however, is to assess whether or not significant differences occur among the different action exemplars, as they did for the typicality ratings. Significant differences among the action exemplars would also further evidence for a radial (prototype) structure for the action categories used in the two experiments.

Each of the 21 subjects responded to each of the 20 conditions for the yes-trials. A 3(Block, between groups) x 4(Action Category, within groups) x 5(Subordinate Exemplars, nested within groups) ANOVA was carried out primarily to test whether the main effects of Block, Action Category and their interaction were significant. If the main effect of Block and the Block x Action Category interaction fail to reach significance, then this would indicate that the different subjects in each block did not lead to different verification reaction times and that the verification reaction times for the different action categories did not vary as a function of the different blocks. There was no reason to suspect that the main effect of block or the interaction between block and action category would be significant.

The reason for using the different blocks had to do with manipulating the no-trials. The analysis of testing for the main effect of block and the interaction between block and action category was performed to rule out these effects and to then allow for collapsing across the block variable in order to perform a per-category analysis. The results showed that the main effects of Block, Action Category and their interaction failed to reach significance [Block, $F(2,18) = 1.10$, n.s.; Action Category, $F < 1$; interaction, $F < 1$]. Given these results, separate one-way repeated measure ANOVAs, collapsed across Block, were performed on the results for the different action exemplars from each category. As in Experiment 1, a Bonferroni adjustment was used to assess the significance of the post-hoc pairwise comparisons at the .005 level.

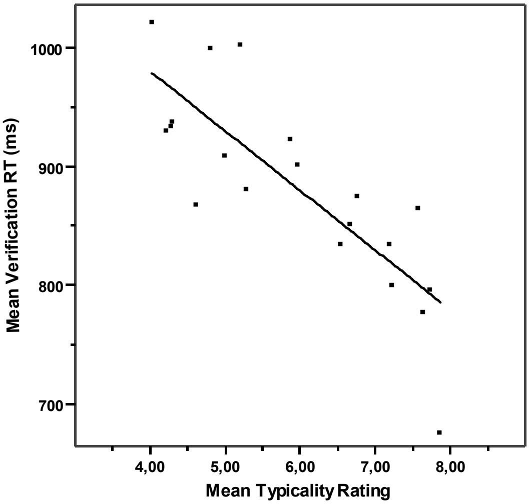


Figure 5.3. Scatterplot for the relation between mean typicality ratings from Exp. 1 and mean verification reaction times from Exp. 2.

Kicking.[36] Results for the KICKING category showed an overall effect between the means for the different action exemplars [*F*(4,80) = 17.73, *MSE* = 9,002, partial η^2 = .470, *p* < .00001]. Pairwise comparisons using a Bonferroni adjustment further showed that SOCCER was verified faster than PUNT [*F*(1,20) = 10.60, *MSE* = 15,220, partial η^2 = .346] and HEEL [*F*(1,20) = 50.92, *MSE* = 20,219, partial η^2 = .718]. PUNT was also verified faster than HEEL [*F*(1,20) = 20.66, *MSE* = 18,183, partial η^2 = .508]. These results are similar to the typicality ratings obtained in the previous experiment where SOCCER was given the highest typicality rating and HEEL was given the lowest typicality rating. The category verification data show that these differences are also significant. The difference between the results for the typicality ratings and the verification times was while PUNT occupied the position between SOCCER and HEEL in the typicality results, KARATE occupied the middle position in the verification results.

---

[36] Reported F-values are based on univariate analyses. The sphericity assumption is fulfilled.

Running.[37] For this category, there were also differences among the means [*F*(4,80) = 9.76, MSE = 35,896, partial η^2 = .328, *p* < .00001]. The only significant pairwise comparisons revealed that SPRINT was faster than BACKWARDS, IN PLACE and SIDEWAYS [*F*(1,20) = 46.05, MSE = 30,220, partial η^2 = .697; *F*(1,20) = 38.84, MSE = 64,464, partial η^2 = .660; *F*(1,20) = 24.28, MSE = 55,087, partial η^2 = .548, respectively]. While SPRINT was verified significantly faster than SKIPPING at the standard significance level of .05, it did not reach significance according to the Bonferroni adjustment [*F*(1,20) = 7.70, MSE = 149,223, partial η^2 = .278]. In contrast to the results for the typicality ratings, the verification results only show a two-tiered relation between the different RUNNING exemplars.

Throwing. The analysis here failed to show any overall difference among the means [*F*(4,80) = 1.97, n.s. Although the verification means show a similar pattern to the means for the typicality ratings, no two-tiered structure could be demonstrated, as was found for the typicality ratings.

Waving.[38] Results for the WAVING exemplars revealed an overall effect [*F*(4,17) = 9.12, partial η^2 = .682, *p* = .0004]. HAND was verified significantly faster than GET BACK [*F*(1,20) = 17.75, *MSE* = 49,780, partial η^2 = .470]. No other comparisons reached significance. A two-tiered structure was also found for the typicality ratings. For the typicality ratings, however, HAND was found to differ significantly from the other WAVING actions, whereas HAND was only found to differ significantly from GET BACK in the category verification task.

While the results from the typicality-RT correlation show a strong relationship between typicality judgments and verification RT, significant differences in verification RT between the different action exemplars for each category only partially confirm the radial structure of the categories. The category for kicking actions showed a three-tiered structure, and verification of different running and waving actions showed a two-tiered structure, whereas the action exemplars comprising the kicking category showed no difference in prototypicality. One important factor that might explain the differences between the results for typicality and verification has to do with the number of subjects. A larger number of subjects would likely lead to less error variance. The standard deviations in Table 5.5 are quite large. The different subjects appear to react differently to the verification task. Increasing the number of subjects or exposing the same number of subjects to multiple

---

[37] Reported F-values are based on univariate analyses. Despite the unfulfilled assumption of sphericity, the different epsilon based corrections do not differ from the results based on the assumption of sphericity.

[38] Since the sphericity assumption was not met, the results from the multivariate analyses (Pillai's trace) are reported here.

runs of the same conditions would likely lead to less error variance. A further discussion of these results will be presented in Chapter 8.

5.2.2.3 No-trials and Context Effect

Analyses of the *no*-trials are intended to assess the effects of functional and perceptual relatedness, i.e., a context effect. Recall that in addition to the yes-trials, where the action exemplars reasonably belonged to the action category signified by the word label, subjects were also presented with *no*-trials in which action exemplars came from contrasting action categories. The results for the *no*-trials are presented in Table 5.7. The means for the related conditions in each block tend to be higher than the means for the unrelated conditions. This suggests that there is some similarity between the action exemplars and the action representation or prototype for the contrast category and that this similarity contributes to the longer reaction times when subjects had to verify that an action exemplar was not a member of the preceding action category.

Block 1 in Table 5.7 contains action labels and action exemplars that can be considered functionally similar. Throwing and kicking actions, although perceptually different body parts are used, are functionally similar in regard to propelling an object. The activation of this information may cause some hesitation when attempting to verify that a specific kicking action is NOT an instance of throwing. In order to assess the effect of relatedness, the two related conditions were combined to form one mean and then compared to the mean of the two means for the unrelated conditions. The ANOVA for used for this analysis, a 2 (related and unrelated) x 10 (action exemplars, nested within relatedness) repeated measures ANOVA, revealed a significant effect of relatedness [$F(1,6) = 11.94$, MSE = 12395, partial $\eta^2 = .666$, $p = .014$].[39] It took more time for subjects to verify the category membership of action exemplars in the related condition (mean = 753 ms) than action exemplars in the unrelated condition (mean = 688 ms). Further interpretation of the results will be presented in the general discussion section below.

Blocks 2 and 3 show a similar pattern of results for the related and unrelated conditions. A 2 (related and unrelated) x 10 (action exemplars, nested within relatedness) repeated measures ANOVA performed on the block 2 data showed that subjects took significantly more time to verify the category membership of action exemplars in the related condition (mean = 981 ms) compared to the unrelated condition (mean = 893 ms) [$F(1,6) = 8.05$, MSE = 28798, partial $\eta^2 = .617$, $p = .036$]. For block 3, however, the difference between the related (mean = 1052 ms) and unrelated (mean = 908 ms) conditions was not significant [$F(1,6) = 2.54$, MSE = 316848, partial $\eta^2 = .298$, $p = .162$]. It is noteworthy that the relatively large

---

[39] Since there are no interpretations of the nested factor or the interaction effect that pertain to the hypotheses being tested here, I will refrain from reporting those results.

difference between the means was not significant in block 3. This is quite likely due to the relatively large standard deviations. Despite that fact that each subject viewed each category label and action exemplar 5 times, the means for each such condition still contained a large amount of error variation.

Table 5.7. Mean reaction time in milliseconds (RT), standard deviation (SD) and total errors (Errors) for related and nonrelated conditions (*no*-trials) for Experiment 2. Blk = Block. Category labels are in capital letters. The arrow indicates the direction of verification. Category labels were presented first, followed by either related or unrelated false exemplar actions.

Blk	Related Conditions	RT	SD	Errors	Unrelated Conditions	RT	SD	Errors
1	THROWING → kicking	744	164	1	WAVING → running	682	107	2
	KICKING → throwing	762	141	10	RUNNING → waving	694	124	2
2	RUNNING → kicking	893	235	9	THROWING → running	847	269	3
	WAVING → throwing	1020	331	17	KICKING → waving	849	246	4
3	KICKING → running	1040	389	14	WAVING → kicking	884	256	3
	THROWING → waving	1063	457	33	RUNNING → throwing	932	322	6

Another general trend is a noticeable difference in the errors for related and unrelated conditions. With the exception of the THROWING → kicking condition, related conditions appear to lead to more errors. Indeed, overall error rates differed significantly, $t(10) = 2.4$, $p = .037$. There also appears to be a linear trend between reaction time and error rate. As subjects take more time to verify the category membership of an action exemplar, they also tend to make more errors; or rather judge those actions as perhaps belonging to contrast categories. The coefficient (Pearson r) for the correlation between reaction time and error rate confirmed the trend, $r = .76$ [$t(11) = 3.72$, $p = .004$].

5.3 General Discussion and Conclusions

The results from the experiments in this chapter support the view that subjects used high-level categorical knowledge in judging the typicality of action exemplars in relation to category labels and when given a speeded category verification task. The results also show an effect of perceptual relatedness indicating access to the spatiotemporal visual shape of actions presented as point-light displays.

Action concepts contain information about characteristic or prototypical spatiotemporal patterns of biological motion. This information can be used to judge the typicality of action exemplars and to make judgments of category verification in relation to previously presented action category labels. Indeed, results from both experiments demonstrate a radial structure for action concepts. Similar to established effects for object categories (Casey, 1992) the results also show a high correlation between judgments of typicality and category verification (typicality-RT effect), which indicates that they rely either on similar processes or on different processes that are similarly affected by the same underlying categorical structure for the investigated action categories. In order to investigate this issue further, it would be necessary to construct an experiment that specifically tested for a dissociation between judgments of typicality and category verification. This is an issue for future research and will not be further investigated here.

The second set of results deals with context effects (Casey & Heath, 1989) where subjects were presented with a category word and an action exemplar that did not belong to the category represented by the word. These ‘false’ trials were investigated in both the typicality rating experiment and the verification experiment. Results from the ‘false’ (non-matching) trials in the typicality rating experiment showed that subjects consistently gave very low typicality ratings to the non-matching exemplars. Exceptions to this pattern occurred for waving exemplars when presented with the throwing label and for throwing exemplars when presented with the waving label. This context effect was also apparent in the data for the non-matching trials from the verification experiment. Verification times for non-matching category-exemplar trials were slower when the category and non-matching exemplars were perceptually similar compared to when they were perceptually dissimilar. This appeared also to be the case for running and throwing exemplars in the context of contrasting category labels, THROWING and RUNNING respectively.

The context effect suggests that action categorization can be affected by the similarity of the spatiotemporal form of an action exemplar to a category prototype. Subjects appear to have access to a prototype representation (template) for a category when presented with a label for that category. It also appears that the prototype representation has a spatiotemporal form such that the perceptual relatedness between the prototype representation of a contrast category label and the viewed action exemplar results in significantly longer reaction times in a speeded verification task.

Regarding the results for the functionally related categories and action exemplars, e.g., KICKING-throwing and THROWING-kicking, in relation to the functionally (and perceptually) unrelated conditions, the fact that category verification took significantly longer for the functionally related conditions than for the unrelated conditions needs qualification. In contrast to the reaction times for the perceptually related conditions, the reaction times for the functionally related conditions were noticeably faster. This was also the case for the unrelated conditions in block 1 as compared to the other unrelated conditions in blocks 2 and 3. Therefore, the significant difference in block 1 should be interpreted with caution, it may be an artifact. The pattern of difference arising between perceptually related and unrelated conditions in blocks 2 and 3, however, indicate a more reliable effect of perceptual relatedness.

There has been little previous research investigating the categorization of actions and the role that action category structure might play in action recognition. Recognizing and understanding the actions of others may be due, at least in part, to having access to action meaning in the form of knowledge about action categories, i.e., groups of similar kinematic patterns of human motion. In his Interactive Encoding Model, Dittrich (1999) suggests that semantic coding might at least involve the ability to make category discriminations. He also equates typicality and related effects as being consistent with the conceptually driven processing of biological motion. Results from the experiments in this chapter indicate that subjects have access to the kinematic templates (prototypes) of various action categories and appear to use that information to make typicality judgments and to verify the category membership of different action exemplars.

Despite very little previous research about the structure of action categories, several researchers have referred to the categorical nature of actions as an explanation for some obtained results. This suggests the need for research on the categorization of actions. The extent of the correlation between RT and typicality was quite large. Together with the context effect, the typicality-RT effect suggests that action and object categories are similarly structured.

The following two chapters address further issues about the properties of action representations and processes involved in the recognition of biological motion. The next chapter more specifically looks at previous experimental findings about the extent to which biological motion perception involves low level or high level processing and the extent to which the visual processing of biological motion is orientation specific. In addition, the role of attention and its interaction with motion orientation will be discussed. The purpose of raising these issues is to gain more knowledge about how humans are able to recognize the actions of others. A further purpose of the next chapter is to provide the reader with the necessary theoretical and empirical background for the experiment in chapter 7.

Chapter 6 - Biological Motion: Levels of Processing and Orientation Specificity

An important feature of the visual processing of the dynamic human gestalt in point light displays is the "automatic" nature of the perceptions. As Johansson (1973) points out, "… we have found that it seems to be a highly mechanical, automatic type of visual data treatment that is most important." While Johansson's use of the term "automatic" points more to the early processes involved in establishing hierarchies of locally rigid perceptual units, there is a case to be made for the automatic processing of biological motion at a higher cognitive level under favorable circumstances, i.e., given an appropriate task. Phenomenally, Johansson's own demonstrations point to the immediateness and vividness of viewing point-light displays of biological motion. Observers are fast and accurate in their identifications when not disrupted by dynamic masking. They appear to have direct access to a level of meaning or semantic level representation that facilitates the identification and recognition of actions depicted in the point-light displays.

Findings from the experiments in the previous chapters support the notion that people have access to and use categorical level knowledge (i.e., how action categories are structured hierarchically, and in relation to an action prototype) when they produce verbs that name natural actions and when they make typicality judgments and engage in category verification. The purpose of this chapter and the next is to further investigate the structure of action categories and the role that categorical information might play in our ability to recognize the actions of others. More specifically, this chapter consists of a literature review relevant to the empirical issues that will be investigated in the next chapter. Consequently, this chapter will describe previous research findings about the extent to which accurate biological motion perception depends on the (vertical) orientation of the point-light display and the level(s) of processing involved. A further issue concerns the role that high-level, or semantic level, knowledge plays in the perception of human actions presented as patch-light displays of biological motion. Could it be the case that high-level knowledge about differences and similarities between different actions is implicitly, or incidentally, involved in simple perceptual tasks that do not require that knowledge? The issue of categorization remains a central topic in this chapter, but it does so within the context of investigating effects of the orientation of the biological motion displays. However, before discussing issues of orientation specific processing, I will discuss distinctions between different kinds of visual processing that are relevant to the perception of biological motion.

The ability to correctly identify the actions of others importantly relies on being able to distinguish between *different* dynamic configurations of the *same* body parts. To what extent is biological motion perception determined by viewing just a few key dots? This kind of processing is also referred to as local motion processing. In contrast, one can also ask what role perceptual access to the whole dynamic display plays in being able to perceive different actions in the point-light displays. This kind of processing is often referred to as global motion processing. Before discussing these alternative processing accounts of biological motion, a discussion of terminology is needed. The terms "global" and "local" have been used to describe different processes involved in being able to see the actions presented in point-light displays of biological motion (e.g., Bertenthal & Pinto, 1994; Thornton, Pinto & Shiffrar, 1998). There appears, however, to be no clear consensus regarding the distinction between global and local processes. To make matters worse, "global" and "local" seem to be synonymous with, or at least not obviously different from, other existing terms. "Global" processing can easily be understood as "holistic" or "configural" processing. Likewise, "local" processing seems similar to analytic, part- or feature-based processing. In addition to these terms, the distinction between levels of processing, i.e., low vs. high-level and bottom-up vs. top down, is likely to co-occur with the global-local distinction.

Firstly, it should be said right away that the different kinds of visual processing of biological motion are task dependent. For example, results from Lange and Lappe (2007) showed that the task of detecting the *facing* direction of a point-light walker required spatial information, not temporal information. However, if the task was to determine the *walking* direction, then both spatial and temporal information were important. It is also likely the case that the task of detecting whether or not a human point-light figure is embedded in a dynamic mask of moving dots requires being able to see the whole human figure or gestalt. On the other hand, being able to distinguish between different kinds of tennis serves (see Chapter 2) requires being able to see local changes in the motion of the arms. There are, however, some helpful attempts at empirically distinguishing between local and global processing, which will be presented below.

6.1 Hierarchical Structure of the Human Body

Attempts to understand the visual processing of biological motion have used a strategy of trying to isolate different sources (levels) of information and then systematically manipulating one source while holding other sources of information constant. The sources included contributions from low-level or local feature based processing on the one hand and more global or configural processing on the other. However, rather than offering a strict dichotomy between local and global processing, Reed, Stone and McGoldrick (2006) (also Reed, Stone, Grubb and McGoldrick, 2006) suggest a *configural processing continuum*. Examples illustrating the different points along the continuum are not restricted to biological motion. At one end of the continuum, we find object recognition which largely relies on the processing of local parts or features. Specific *local object features*, e.g., parts, can be largely diagnostic for identifying objects like houses and cars (Biederman, 1987; Gauthier & Tarr, 2002). At this far end of the continuum, there is very little, if any, configural processing (cf. Boucart & Humphreys, 1992). At the other end of the continuum, processing can be characterized as completely *holistic*, and thereby configural, in the sense that processing is based on access to *template-like unparsed wholes*, i.e., as in face recognition (e.g., Farah, Wilson, Drain and Tanaka, 1998).

Between the two endpoints on the configural processing continuum, varying degrees of configural processing can occur. The processing of *first-order information*, for example, involves determining the basic spatial relations between relevant parts. Let us take the human face as an example to illustrate first-order information. The chin is *below* the mouth, the mouth is *below* the nose, and the eyes are located slightly *above* the nose. Reed, Stone and McGoldrick (2006) suggest further that *structural information* about the face includes not just first-order information but first-order information as it relates to the *hierarchical organization* of object parts in relation to the *overall* object. Structural hierarchical information, therefore, is viewed as more configural than first-order information. Another kind of visual processing is based on *second-order relational information*. In the case of face recognition, second-order relational information refers to specific relative distances between the hierarchically organized different parts of the face, i.e., the distance between the eyes and the distance between the nose and the mouth. Humans seem to be sensitive to changes in the second-order relational information in human faces. (For extended discussions of configural processing in face recognition see for example Boutsen and Humphreys, 2003; Calder and Jansen, 2005; Carey and Diamond, 1994; Diamond and Carey, 1986; Leder and Carbon, 2006; Maurer, le Grand and Mondloch, 2002.)

Similar to the human face, perception of the human body can also be characterized in terms of different kinds configural processing. Marr and Nishihara (1978), for example, described the structural hierarchy of the human body in their 3D model representation. According to this model, the human body (and other animal bodies) can be represented according to the relative arrangement of the component axes of body parts (first-order information) in relation to the principle axis of the human body as a whole (structural hierarchy). The following quote from Bosbach, Knoblich, Reed, Cole and Prinz (2006, p. 2951) relates previous findings from experiments on the perception of body postures to the Marr and Nishihara idea:

> The processing of configural relations of a body posture seems to rely on the structural hierarchy of body parts, not on the isolated parts themselves. Structural hierarchy refers to the organization of isolated body parts in terms of the overall object and the spatial relationship of each part relative to each other (cf. Marr, 1982). For instance, bodies are recognized not only by the fact that the shoulder and arms are below the head but also from the fact that the shoulder and arms are in a particular position relative to the overall structure of the body, that is, they are always attached to the same part of the trunk and above the feet. (p. 2951)

The findings from Reed, Stone, Grubb et al. (2006) support this idea of a body representation or body schema that is described as a hierarchical topological representation. The body representation occupies a place between first-order configural processing and the holistic processing end of the continuum. Evidence strongly suggests that configural processing is central to both face recognition and the recognition of human body postures (Reed, Stone, Bozova & Tanaka, 2003; Reed, Stone & McGoldrick, 2006).

Local body-part motion alone is not sufficient to create the impression of human body form. Limb symmetry and other information that indicates the hierarchical organization of the human body around a principle axis of the body torso seems to be critical for, at least, recognizing the form of the human body. For recognizing specific actions, a possibly more demanding task, or at least a more specific task, information about the relative nested motion of hierarchically organized body parts may be needed. Reed et al. (2006) maintain that the visual recognition of static body postures and dynamic bodies share similar demands on configural information. A viewer needs information about the hierarchical organization of the human body. (See Casile and Giese (2005) and Thurman and Grossman (2008) for results showing an exception to this.)

While acknowledging the lack of clarity concerning the distinction between local and global processing, Pinto and Shiffrar (1999), for example, offer a description that captures important aspects of the distinction. Local processes/mechanisms are limited in the features that figure in the processing. They are also limited in respect to the spatial extent over which information is integrated, and they take place relatively early on in visual processing. Global processing, on the other hand, is more sensitive to changes that occur over a larger spatial extent, perhaps on the level of a delineated whole object. Feature changes that affect the interpretation of a whole object will be registered by global processes/mechanisms. In the case of biological motion perception, the different nested motion of body parts give can rise to distinctly different actions such as walking or climbing. More local processes might be needed to detect minor differences in the movement of an arm or hand in the case of distinguishing between different ways of throwing or waving. Pinto and Shiffrar

(1999) suggest that the recognition of human actions in point-light displays is an example of category-specific processing.

6.2 Role of Form and Motion Cues

Displays of biological motion differ from both static displays of faces and body postures by the addition of a critical temporal, or dynamic, component. Point-light displays of biological motion also differ from face and body posture recognition by the fact that they contain relatively little form information. There is no explicit information about the contour of the human body. This, however, does not prevent subjects from possibly extracting body structure from the motion of the dots. The visual processing of point-light displays requires the (simultaneous) tracking of points over time where the points of light have different velocities and are connected in a way that strongly suggests a coherent whole. There is no visible body contour by which to perceptually group the moving points of light.

Reed, Stone and McGoldrick (2006, p. 244)) state that “A configural representation of the body and how its parts relate to the whole-body hierarchy in terms of structure and biomechanics constrains the interpretations of the point-light movements.” Here, they introduce the notion of a global analysis. What is the nature of a global analysis in relation to configural processing or a configural representation? They mention (p. 244) that “the visual system performs a global integration of motion signals over time and space to create a representation of body configuration.” In biological motion perception, this means that there are some visual processes that are sensitive to spatial and temporal configural information of the human body.

Pyles, Garcia, Hoffman & Grossman (2007) discuss their findings in relation to the possible role of dynamic templates in action recognition. They constructed dynamic ‘Creatures’ that had bodies with articulated joints and could therefore move within their environment. One critical difference between Creatures and human bodies was the fact that different Creatures had *different* body configurations, while different humans have the *same* body configuration. While the human body configuration and its various motions are very familiar to us, the Creature configuration and motion is unfamiliar. Subjects viewed fully illuminated and point-light versions of the movement of the Creatures and some familiar actions performed by human body, e.g., kicking, jumping and throwing. Two results are of interest here. Firstly, when point-light Creature movements and familiar human actions were masked with scrambled motion, subjects showed little tolerance for the scrambled motion when it occurred together with Creature movements. Subjects were more tolerant of the scrambled motion when it occurred together with the familiar human actions. This indicates that access to high-level, global information is severely impaired for the Creature movements compared to the human actions.

Recent research has attempted to determine the respective roles that form and motion information play in biological motion perception. The importance of form, or body-based, information has been emphasized in the findings of, e.g., Lange and Lappe (2006) and Beintema and Lappe (2002). In contrast to this view, others have asserted the critical role that local motion plays in the perception of biological motion (e.g., Casile & Giese, 2005; Giese, 2006; Giese & Poggio, 2003; Thurman & Grossman, 2008). In order to clarify the potential respective roles of form- and motion-based information, I will briefly describe these two approaches. (For further details, see Giese (2006) for a thorough overview of the different modeling approaches used in motion recognition.)

Beintema and Lappe (2002) created a point-light walker in which the dots were randomly positioned for each frame of the point-light sequence between the joints of the body instead of being placed directly on the joints. This effectively ruled out any kind of processing that attempted to track the motion of an individual point-light (local motion) to gain information about limb movement, i.e., *form-from-motion*. It did *not*, however, rule out visual processing that could use the point-light positions associated with each sequence frame to create a form template, which could then be temporally integrated to create a *motion-from-form* dynamic template. Even on the basis of the impoverished motion displays, subjects could accurately discriminate the direction of articulation of a point-light walker. As further support for the primary role of form-based information, Lange and Lappe (2006, 2007) also created a neurally plausible model based on their previous findings as well as the findings of others. The model uses template matching of global, configural form information and explicitly has as one of its assumptions that "biological motion may be inferred from form analysis *without* local motion processing" (Lange & Lappe, 2006; p. 2896, italics added). The performance of the model matched data obtained from psychophysical and neuroscientific experiments (Lange, Georg & Lappe, 2006; Hirai & Hiraki, 2006). These results indicate that biological motion perception can be accomplished by global form analysis.

The previously mentioned model (chapter 2) of Giese and Poggio (2003) differs from Lange and Lappe's model to the extent that it emphasized the role of both form- and motion-based information, and it is also hierarchical, i.e., recognition is built up from a hierarchy of neural feature detectors. Most important, however, is the finding that subjects could reliably identify a human point-light walker on the basis of the opponent motion of the wrists and ankles (Casile & Giese, 2005; Troje & Westhoff, 2006). This was demonstrated by constructing a critical features stimulus that only included local opponent motion, i.e., the sinusoidal and antiphase components of the wrists and ankles. Motion vectors for the other dots were completely random. Casile and Giese (2005) also showed that opponent motion information is similar for both normal (whole body) and point-light stimuli, which might explain to some extent why the recognition of actions using whole body stimuli generalizes to point-light displays.

According to the recent findings of Thurman and Grossman (2008), local opponent motion of the extremities, i.e., wrists and ankles, is most important for discriminating point-light displays of biological motion. This was in contrast to velocity considerations and apparent body structure. Discriminating jumping jacks from walking appears to rely crucially on the opponent motion of, for example, the ankles. The fact that these findings suggest an important role for local information calls into question the role that configural information might still play in the perception of biological motion. However, Thurman and Grossman (2008) state the following (p. 9): " It has been suggested that highly familiar complex patterns such as this could make up a "vocabulary" of sorts, for which we may develop dynamic templates (Cavanagh, Labianca & Thornton, 2001). … Biological motion would therefore be an ideal candidate for these putative dynamic templates." In other words, the findings of a critical role for local opponent motion do not rule out a role for more configural processing.

Thurman and Grossman explicitly assert that *body posture* as such is unlikely to play a critical role in discriminating point-light animations. In my opinion this should be interpreted to mean that *static* body posture with no motion information is unlikely to play a critical role. So, in addition to the findings of a critical role for local motion opponent information in discrimination tasks, it could be the case that some action representations in the form of dynamic templates may be involved in more fine-grained visual distinctions between, for example, different kicking actions. Casile and Giese (2005) state that other, perhaps more sophisticated, tasks like distinguishing between the genders of point-light walkers or determining the emotional content in a point-light may require additional information. They also suggest that there is likely a role to be played by higher level cognitive representations. For example, Giese and Poggio (2003) also discuss the notion of prototypes. In their model, representations are stored as two-dimensional prototypical patterns of human movement. Their model is also consistent with data obtained from psychophysical and neuroscientific experiments.

Although it appears that the two sides of this debate agree on the question (What is the relative importance of motion- and form-based information in the visual processing of point-light displays of biological motion?), they disagree about the answer. Whereas Lange and Lappe (2006) emphasize the role of global form analysis, Casile and Giese (2005) emphasize the role of local opponent motion of the extremities. Could it be that both sides are to some degree correct? Blake and Shiffrar (2007) draw the conclusion that both form and motion have been demonstrated to play important roles in the perception of human action. Findings from neuroscience also indicate the integration of motion and form information in perceiving the actions of others. For example, numerous studies have shown consistent activation of the posterior superior temporal sulcus (pSTS) when observing point-light displays of meaningful actions. There is general consensus that the pSTS is particularly involved in the integration of motion and form. There are of course other brain areas involved in the visual processing of biological motion, e.g., premotor cortex (Saygin et al., 2004). See Grossman (2006) for a review of the area. The relation to neuroscience will be described in a later section.

6.3 Temporal Aspects of the Action Representation

Other researchers have also considered the role that more configural dynamic processing plays in biological motion perception. Access to global motion patterns has also been described in terms of "sprites" (Cavanagh, Labianca & Thonton, 2001). Sprites are attention-based visual routines used to recognize familiar motion patterns. Given the familiarity of the actions portrayed in studies of biological motion, action recognition seems to be facilitated by a stored motion pattern that matches the global motion pattern of the input. This suggests that action recognition is mediated in a top-down fashion by dynamic representations of the human form. Research on representational momentum also supports this claim by showing that subjects represent the possible dynamic paths of human motion in priming studies (e.g., Freyd, 1983, 1987; Kourtzi & Shiffrar, 1999; Verfaillie & Daems, 2002). Consistent with the behavioral studies on representational momentum, Kourtzi and Kanwisher (2000) also showed that there was significant cortical activity in motion sensitive areas in the brain when subjects viewed a static image of an athlete getting ready to throw a discus. This indicates that motion information is stored and activated together with structural properties of the visual stimulus.

As further evidence of the dynamic component of the action representation, Verfaillie, De Troy and Van Rensbergen (1994) sought to determine the extent to which changes in biological motion displays were detected across saccadic eye movements. Since visual information is largely suppressed during saccadic eye movements, some presaccadic information has to be retained in order to maintain a coherent experience of motion across saccades. Verfaillie et al. (1994) showed that subjects were relatively poor at detecting postsaccadic forward shifts of a point-light walker when those shifts occurred during a saccade and the shifts were consistent with the expected forward trajectory of the point-light walker. However, subjects were better at detecting backward shifts under the same conditions. Verfaillie et al. (1994) reason that the visual system can anticipate the step-cycle position of the point-light walker across saccades. So, when the anticipated step-cycle position is *consistent* with presaccadic forward shifts, they were more *difficult* to detect because they were expected. Backward shifts were *unexpected* and therefore *easier* to detect.

The temporal aspects of biological motion perception are not restricted to the visible stimulus, but rather extend past the endpoint of the visual stimulus to include information about the movement of the body as if it were to continue in motion. In another series of experiments that investigated the anticipatory visual processing of actions, Verfaillie and Daems (2002) let subjects view whole-body (i.e., not point-light) animations of human actions during a priming phase. In one condition following the priming phase, subjects were presented with a static body posture that was a continuation of the animation in the priming phase, but was not previously seen. The task was then to indicate whether or not the static body posture was possible or impossible. The results showed that subjects responded faster when the static body posture was a continuation of the animation previously viewed in the priming phase compared to a static body posture that was not previously primed. The results from this study support the hypothesis that human action perception engages anticipatory processing which in turn appears to rely on the temporal aspects of the action representation.

Recent results from Graf et al. (2007) also indicate that temporal aspects, i.e., future body posture, are important features of the action representation. Indeed, it appears as if the motor system also informs perception about the changes in the relative position of body parts as a given action is performed or observed. Our ability to perform certain actions allows us to simulate the actions in real-time. Graf et al. (2007) let subjects view 9 different actions (e.g., throwing a ball, waving both arms, lifting something from the floor, etc.) presented as point-light displays. Critical to the experiment was the fact that a portion of the point-light sequence was occluded. There were 3 different occluder durations, 100 ms, 400 ms and 700 ms. In addition, there were also 3 different movement gaps, also 100 ms, 400 ms and 700 ms. This allowed for the independent manipulation of occluder duration *and* movement gap. Normally if a movement is occluded, the duration of the occlusion matches the portion of the movement that was occluded, and the first frame of the sequence following the occlusion contains a posture that the viewer would normally see as a natural continuation of the action sequence immediately following the occlusion. By independently manipulating occlusion duration and movement gap, it was possible to occlude a point-light sequence for 100 ms and then show the position of the body as it would have been if 700 ms had passed instead of 100 ms. In this case, the occlusion is too short and the sequence appears to jump ahead in time. If action perception involves real-time simulations of observed actions, then subjects should make the fewest errors when occlusion duration and movement gap are in agreement, i.e., both are 100 ms, 400 ms or 700 ms. The task for the subjects was to decide whether the posture of the point-light actor immediately following the occlusion was a continuation of the action in the same orientation or a different orientation. The results showed that subjects indeed made the fewest errors when occlusion duration and movement gap were in agreement, and there was a significant linear trend showing that error rate increased with increasing distance between occlusion duration and movement gap.

People appear to predict the immediate future body postures of familiar actions. This finding of reliable action prediction supports the role of action representations in real-time simulation of external events. (See also Wilson (2006) for an extended discussion of the role of a human body emulator in covert imitation.) Further evidence of the internalized temporal, or dynamic, nature of the action representation comes from experiments on motor imagery. Jeannrod (2006) for example discusses the occurrence of Fitt's law in mentally executed movements. Fitt's law states that the time it takes to point to a target object is a function of target size and distance. Subjects apparently internalize the interaction of physical constraints in the environment and the physical constraints of the human body. Many of the physical constraints that shape the visual appearance and the internal representation of human movements have also been documented using point-light displays. Runeson and Frykholm (1983) for example, demonstrated that subjects are sensitive to the dynamics that constrain the motion patterns in the point-light displays. Subjects in their studies could reliably detect the various weights of an unseen box lifted by the point-light actor. The length of a throw could also be determined on the basis of the pattern of motion of the thrower. Even the gender of adults and children were detected with reliable accuracy. More remarkable was the finding that subjects were able to "see" deception. When the point-light actor was instructed to try and make a lifted box look heavier than it actually was, subjects could see that this was the case and therefore could perceive the intention to deceive. They were also able to estimate the faked weight of the box.

Within the context of the discussion about the configural processing of biological motion, a question arises as to how to characterize the temporal aspect of configural processing. What kind of dynamic information could be considered as configural information along the temporal dimension? Is there such a thing as dynamic configural processing/information? It could be the case that distinctions between first-, second-, and third-order motion indicate different levels of configural processing for displays of motion. Garcia and Grossman (2008) investigated the role of different levels of motion processing in biological motion perception. They specifically looked at the role that first-, second-, and third-order motion information play. While first-order motion can be obtained by small changes in luminance across space, second-order motion requires differences in relative contrast or texture across space. So, even when two areas have the same *average* luminance, motion can be perceived if there are sufficiently large contrast changes between the areas. Our ability to track third-order motion depends, however, on more high-level contrast perception. In this case, third-order motion mechanisms use attention to track contrast dependent movement. According to Garcia and Grossman (2008), it is this kind of motion perception that is necessary but not sufficient for biological motion perception. Access to some kind of form information, i.e., structure from motion, also seems to be required, and this in turn appears to require attention.

6.4 Configural Processing and the Inversion Effect

When unmasked displays of biological motion are viewed in an upright orientation, the depicted actions are easily recognized. They seem to pop out. There is phenomenally direct access to a global, semantic level of representation. An effective method of disturbing the categorical/semantic processing of biological motion is to turn them upside down, i.e., invert them. There is a wealth of converging behavioral and neuroscientific results that demonstrate impaired recognition, identification, detection and priming when displays of biological motion are viewed upside down (inverted) (e.g., Ahlström et al., 1997; Bertenthal & Pinto, 1994; Daems & Verfaillie, 1999; Dittrich, 1993; Grossman & Blake, 2001; Pavlova & Sokolov, 2000; Shiffrar & Pinto, 2002; Sumi, 1984; Troje, 2003). The effect of changing the viewing orientation for point-light displays can be defined as a comparative difficulty in recognizing, identifying or detecting point-light displays of human actions that differ from a previously viewed orientation or differ from some established canonical orientation. In addition to this inversion effect, other studies have demonstrated orientation specificity in-depth. Using short-term priming, Verfallie (1993, 2000) found significantly more priming for congruent in-depth displays (right and left facing point-light walkers) than for incongruent displays that differed in their in-depth orientation. Olofsson, Nyberg and Nilsson, (1997) obtained similar results using a long-term priming paradigm. Subjects were significantly better at naming previously seen displays depicting various actions if the primed action had the same left-right in-depth orientation. Long- and short-term priming appear to lead to similar results of view dependence for in-depth orientations.

In contrast to in-depth manipulations of point-light displays, where separately left-facing and right facing displays are easily recognized, when a display is inverted, subjects have difficulty recognizing or naming the action depicted in the display. The theoretical significance of this inversion effect has to do with the fact that inverted displays contain the same hierarchical structural information as upright displays. The same local pair-wise relations and their relations to a principal axis of organization occur in inverted and upright displays. The performance differences between perceiving upright and inverted displays indicate different processing mechanisms, and therefore by systematically investigating performance differences under varying experimental conditions, we may gain further insight into our understanding of the factors that influence our keen ability to perceive the actions of others. A crucial step in understanding the ability of humans to quickly perceive the actions depicted in point-light displays is finding out under what conditions this ability fails, and the inversion effect has functioned as kind of benchmark for demonstrating difference between local and global/configural processing. Global processing seems to be critical to perceiving upright displays, and the inversion effect occurs due to a lack of global processing. (See, however, Chang and Troje (2008), who show that local motion processing can also give rise to the inversion effect.) The following results from various experiments demonstrate the robustness of this effect and the conclusions that can be drawn from it about differences in visual processing.

In one of earliest studies demonstrating the inversion effect, Sumi (1984) let subjects view inverted walking and running sequences. The majority of subjects who reported seeing a human figure failed to see it as inverted. They reported arm movements for the legs and leg movements for the arms. Other responses included non-human elastic forms indicative of non-rigid motion and mechanical changes. These results suggest that human perception of biological motion is sensitive to the image plane orientation of the displays. Of particular importance is the fact that subjects apparently were able to see local motion in terms of the motion of arms or legs but failed to get "the whole gestalt." A number of studies since then have systematically investigated this phenomenon.

Behavioral data from Dittrich (1993) showed a clear inversion effect for most of the actions included in that study. Subjects' ability to recognize the different actions was severely impaired when the actions were shown up-side down. However, it was not the case that recognition was completely disrupted. Dancing, for example, had a recognition rate of 61% when it was view up-side down compared to 87% in the upright condition. Using a detection task, Bertenthal and Pinto (1994) showed that subjects were unable to detect an inverted point-light walker when it was embedded in a mask of randomly positioned dots that had the same local trajectories as the dots in the target point-light walker. For upright walkers, detection was significantly above chance. In this case, when the local motion trajectories of the point-light walker were duplicated by the randomly placed masking elements, local motion processing was effectively prevented. Bertenthal and Pinto (1994) reasoned that if subjects, however, were still able to detect the point-light target when it was dynamically masked, then it would follow logically that visual processing occurred on a more global level.

In order to more systematically investigate the extent to which the perceptual stability of point-light displays is affected by rotation in the image plane, Pavlova and Sokolov (2000, exp. 1) let subjects view a non-masked point-light walker. The walker was presented first as inverted and then incrementally rotated 30 degrees until it was completely upright. Subjects were instructed to indicate when they experienced a change in interpretation of the display. If a perceptual stimulus gives rise to many different interpretations, then the stimulus is unstable relative to a stimulus that does not give rise to many different interpretations. The results from this experiment showed that a 90 degree orientation for the point-light walker led to the most interpretations. When the display was completely inverted, i.e., 180 degrees, there were still relatively many interpretations relative to when the display was presented upright. In addition to this finding, Pavlova and Sokolov (2002, exp. 4) used a long-term priming paradigm to investigate possible priming effects as a function of different orientations. They found only priming effects for upright or near to upright (0-45°) point-light walkers, where an upright prime display had a priming effect on both identical upright walkers and on walkers rotated 45° from upright, suggesting that a "… priming effect in biological motion is *partly* independent of the relative orientation of priming and primed displays" (Pavlova & Sokolov, 2000, p. 897). There was no pronounced priming effect for masked congruent inverted displays, i.e., a masked inverted point-light walker did not prime a masked inverted walker in a detection task. Further findings from Pavlova and Sokolov (2003) indicated that informing subjects about the inverted orientation of soon-to-be-presented stimuli did not lead to better recognition performance when compared to subjects who did not receive that information. The results suggest that the inversion effect can not be diminished through the top-down influence of prior information about the orientation of the stimuli.

In an investigation of the visual bistability of point-light walkers, Vanrie, Dekeyser and Verfaillie (2004) found that subjects were much more biased towards seeing an *upright* point-light walker as being oriented towards the viewer, i.e., the subjects, than oriented away from the viewer. This occurred despite the fact that no explicit depth cues were available to specify a unique visual interpretation. However, for the *inverted* point-light walker, no such visual bias occurred. Subjects reported seeing a point-light walker facing the viewer about 50% of the time. This visual bias difference between upright and inverted point-light walkers suggests that in relation to an upright walker, the inverted walker disrupts the processing such that no stable perceptual interpretation about the facing direction of the walker occurred. Although Vanrie et al. (2004) stop short of describing the mechanisms that explain the results, they seem to suggest that global processing occurs to a larger extent for upright than for inverted point-light walkers. Thornton, Vuong and Bülthoff (2003) suggest that complex motion perception, resolving perceptual bistability, may be mediated by top-down access to stored dynamic templates. The role of the templates is to understand the meaning of the stimulus within the context of previous experience and knowledge.

Results from Loula et al. (2005) show that people are more sensitive to their own movements than the movements of strangers. Expressive actions like boxing and dancing carry specific information about the individual performing the action, and are subsequently recognize better than actions that are more “common” like running and walking. The recognition of actions in inverted displays is not greater than chance and shows that there is little, if any, processing of configural information in the inverted displays.

Ikeda, Blake and Watanabe (2005) provide further evidence for the role of configural processing in perceiving upright and inverted biological motion. They found that our ability to distinguish coherent displays of biological motion (kicking, jumping, walking, running and throwing a ball) from scrambled displays depends on where the biological motion stimulus appears in the visual field. Visual processing for upright displays was very sensitive to changes in stimulus position away from foveal vision, despite compensating changes in spatial scaling. For example, changes in 4 degrees of eccentricity from central vision led to a relatively large decrease in performance. It was also found that the inversion effect depended on stimulus eccentricity. The ability to distinguish between coherent and scrambled biological motion displays was better for upright displays compared to inverted displays for foveal vision. However, when the upright and inverted displays were presented at the 12 degree periphery, there was no such advantage for the upright displays. For this condition, there was no inversion effect. This, however, was not due to the fact that subjects had a difficult time seeing the motion of the dots. Ikeda et al. (2005) report that subjects were able to see dot motion in nearly all of the conditions. Instead, it appears that visual processing of biological motion in the periphery lacks the capacity to perceptually group the moving dots into coherent meaningful percepts. They also conclude, somewhat speculatively, that the results point towards the use of an active top-down processing strategy that has access to stored representations of different kinds of actions.

Recent studies using normal control subjects and subjects with amblyopia[40] have been carried out to determine the extent to which there may be processing differences for upright and inverted point-light displays. In one of these studies, Neri, Luu and Levi (2007) investigated the extent to which global form processing might be reduced in amblyopic subjects. In order to isolate the global form-based processing, Neri et al. (2007) rendered the local motion trajectories of the point-light display uninformative by masking them with dots that had the same trajectories. Subjects were then given the task of discriminating between a target (a coherent point-light action) and a non-target (a scrambled version of the coherent point-light action). This ensured that the only difference between the targets and non-targets was the global form information in the targets. Targets and non-targets were also presented either upright or inverted. Subjects separately viewed the stimuli with their amblyopic eye and their non-amblyopic eye. The results showed that the inversion effect occurred for the amblyopic eyes as well as for the non-amblyopic eyes, even though performance was diminished for both upright and inverted displays for the amblyopic eyes. On the basis of these results, Neri et al. (2007) maintain that visual processing in amblyopic subjects includes intact global form-based processing. They also use the term ‘high-level’ to refer to later stages in the motion processing hierarchy. They suggest, in line with previous descriptions, that ‘high-level’ refers to processing that attempts to retrieve structure from motion. The term is intended in a broad sense to include motion processing that takes place after optic flow extraction and after the processing of local information occurring in translation and rotation patterns. They conclude that access to stored ‘high-level’ upright motion patterns is preserved in amblyopic subjects. Inverting the point-light displays disrupts this final recognition stage of visual motion processing.

---

[40] Thompson et al. (2008, p. 1) describe amblyopia as “a developmental disorder of the visual system caused by ocular abnormalities early in life. While surgery or optical correction of refractive errors can often address the initial cause of the amblyopia (e.g., strabismus), once amblyopia has developed, such interventions cannot restore visual function since amblyopia itself is a cortical deficit.” Furthermore, they add that visual processing deficits occur in striate and extra striate areas.

The inversion effect might not, however, be limited to the lack of global form-based processing in inverted displays. Thompson, Troje, Hansen & Hess (2008) suggest that there might be at least two separate inversion effects, one that is the result of visual access to the global form-based information in the point-light stimulus and the other that is the result of visual access to local motion information. In order to test this idea, Thompson et al. (2008) created two sets of biological motion displays. One display preserved the global form-based information by masking local motion information of a coherent point-light walker with a dynamic mask, and the other display preserved local motion information by scrambling the otherwise coherent global pattern of the dots but yet retaining the motion trajectories of each individual dot. Thompson et al. (2008) let amblyopic subjects view the different displays and try to discriminate the walking direction of the coherent (preserved global form) and scrambled (preserved local motion) point-light walkers. The results confirmed the previous findings from Neri et al. (2007) but also showed that the inversion effect could be obtained for the local motion displays for both amblyopic and control subjects. The inversion effect, however, was larger for the scrambled displays in the amblyopic subjects compared to the control subjects, suggesting that amblyopic eyes have more difficulty seeing the global form in the coherent display. Thompson et al. (2008) view their results as supporting previous findings (e.g., Casile & Giese, 2005; Troje & Westhoff, 2006) showing that local motion information can play a role in biological motion processing.

6.4.1 Visual Learning and the Inversion Effect

Although it is certainly the case that the visual perception of point-light stimuli is disrupted when they are inverted, some results show that subjects are able to learn to visually process inverted stimuli successfully under some conditions. For example, the results from the studies of Pinto and Shiffrar (1999) showed that detection performance for inverted displays, although impaired, was significantly better than chance. Furthermore, in a series of experiments that addressed the issue of learning arbitrary and biological motion, Hiris, Krebeck, Edmonds and Stout (2005) found that significant learning could occur for inverted biological motion stimuli, but learning was limited to a specific circumstance. When subjects were given the task of detecting inverted stimuli, and detection could be successfully performed by only focusing on the motion of a few dots, the results showed that the detection of inverted stimuli reached performance levels that were similar to the upright stimuli. Critically, however, when subjects were required to use a more global processing strategy in order to determine the whether or not the bottom *and* top halves of the point-light target were moving coherently, performance for the inverted stimuli was severely reduced to levels similar to the visual processing of arbitrary stimuli. These results are consistent with the previously mentioned studies that suggest that global processing is severely impaired for the inverted point-light displays.

Relevant to the issue of learning of biological motion, Jastorff, Kourtzi and Giese (2006, exp. 2) let subjects view 2 kinds of point-light displays. One kind of display consisted of movements that were easily perceived as being human-like. These movements were created by morphing between two “real” actions, e.g., running and marching. The other kind of display consisted of non-human-like artificial movements. The movements, however, were coarsely matched to the structure of the human-like displays. The purpose of the experiment was to investigate the extent to which learning human-like and non-human-like movements may differ. It could be the case that learning mechanisms may be more sensitive to human-like stimuli than non-human-like stimuli. Learning was assessed by interleaving training phases with subsequent test phases. In the test phases, subjects had to state whether or not a second stimulus matched the first one. A further manipulation in the experiment was the orientation of the stimuli during training and testing. During training, the stimuli were rotated by 90 degrees (i.e., not complete inversion) in the image plane. The stimuli in the test phases were shown in the “usual” upright orientation. The issue addressed with this manipulation was the extent to which discrimination performance would “carry over” from the rotated training phase to the upright test phases. The results showed that while subjects were able to learn the rotated stimuli, there was no performance generalization from the rotated stimuli in the training phase to the upright displays in the test phase. This was true for both the human-like and non-human-like stimuli. The visual experience of a point-light movement from one perspective or orientation does not seem to facilitate the visual processing of a point-light movement from another perspective. Jastorff et al. (2006) also showed that fast discrimination learning for articulated point-light movements depended on access to information about the underlying skeleton of the human actor. Discrimination was much more difficult when that information was spatially scrambled and therefore not available to visual processing.

6.4.2 Brain Activity and the Inversion Effect

Recent data using fMRI shows that the articulated movement of human bodies leads to significant activation of the posterior superior temporal sulcus (pSTS) and to its important role in biological motion perception (e.g., Grossman, Battelli & Pascaul-Leone, 2005; Peelen, Wiggett, & Downing, 2006; Peuskens, Vanrie, Verfaillie & Orban, 2005; Pyles et al., 2007; Vaina & Gross, 2004; Vaina, Solomon, Chowdhury, Sinha & Belliveau, 2001). When it comes to processing upright point-light displays of biological motion, the pSTS (largely right hemisphere, but also bilateral) seems to be particularly critical. For example, Peelen et al. (2006) let subjects view 4 kinds of stimuli: solid whole body figures, point-light actions, faces and scrambled point-light controls. One of the important contributions in this study concerned the dissociation between different posterior cortical areas that become activated by point-light displays. While the extrastriate body area (EBA) and the ‘fusiform body area’ (FBA), appear to respond to the presence of the *form* of the human body, the pSTS appears to integrate body *movement* information over time. Pyles et al. (2007) also obtained fMRI data while subjects viewed fully illuminated and point-light "Creature" movements and human actions. "Creatures" had articulated joints connected to a body which allowed them to move through their environment. The advantage of creating these kinds of creatures with body configurations that were piece-wise rigid was that point-light animations of Creature movement could be created in a way similar to human point-light displays. The major difference between Creature and human movement was the different body structures. Blood oxygenation level dependent (BOLD) signal changes were measured when subjects viewed fully illuminated and point-light stimuli. The results showed a significantly higher level of activation for the familiar human actions than for the Creature movements, even when subjects had previously given the Creature movements high ratings of animacy. It appears that the pSTS is selectively sensitive to familiar meaningful human action. These results support the previously mentioned hypotheses regarding the role of high level dynamic motion templates that represent the underlying structure and predictable motion of the ways in which humans move. Given previous results showing that inverted biological motion displays are more difficult to identify, recognize and detect under various circumstances, what role might processing in the pSTS play in our ability to perceive the actions in inverted point-light displays?

In an fMRI study, Grossman and Blake (2001) had subjects view upright, inverted point-light displays depicting various actions. Scrambled point-light displays were also viewed as a baseline condition. The point-light displays were not masked. A 1-back task was used during the scanning sequences. For this task, subjects were to respond when the current stimulus was identical to the immediately preceding one. The results showed significant BOLD responses in 6 of their 8 subjects for the inverted point-light displays in pSTS compared to the scrambled displays. Activation in pSTS was also significantly greater for upright displays than for inverted and scrambled displays. These results suggest that there is some level of visual processing in the pSTS that supports the perception of inverted displays. Verbal reports from the subjects in Grossman and Blake (2001) indicate that some point-light actions were recognized as being presented upside-down. Given the fact that a critical difference between the scrambled and inverted displays had to do with the absence or presence of a coherent form, it seems reasonable to claim that some of the processing of the inverted displays was related to the coherent form of the inverted displays. So, there may be enough global form processing to support the level of performance observed for the inverted displays in the 1-back task, but not enough to reach the level of performance observed for the upright displays.

In a further attempt to understand the role that pSTS plays in the perception of biological motion, Grossman et al. (2005) used repetitive transcranial magnetic stimulation (rTMS) to temporarily disturb processing within the pSTS in the right hemisphere and then measure the performance of biological motion processing. Subjects in this experiment viewed different point-light actions, i.e., kicking, walking, throwing etc. The point-light actions were presented upright and inverted. Scrambled versions of the actions were also created and presented to the subjects. All stimuli were embedded in a dynamic mask to reduce the occurrence of ceiling effects. It should be noted here that there were fewer masking elements for the inverted displays since the perception of inverted displays is more difficult. The task for the subjects was to indicate which displays contained the non-scrambled and scrambled point-light actions. The results showed that performance *prior* to the administration of rTMS was *similar* for upright and inverted displays, which was largely due to the fact that the inverted displays contained fewer masking elements. When, however, performance was measured during the time frame for processing disruption due to rTMS over the pSTS, performance *decreased* significantly for the upright displays compared to the results prior to administration of rTMS. For the inverted displays, there was no such reduction in performance. It should also be mentioned that rTMS applied to area MT+/V5 in the left hemisphere, another motion sensitive area for more low level motion, *did not* have an effect on performance for upright or inverted displays. Processing in the pSTS is apparently necessary for the configural processing of upright displays. For the inverted displays, there appear to be some other processing mechanisms that are either outside the pSTS or were not sufficiently disrupted by the applied magnetic stimulation.

Using magnetoencephalography (MEG), Pavlova, Lutzenberger, Sokolov and Birbaumer, (2004) had subjects perform a 1-back repetition task when viewing an unmasked upright and inverted point-light walker. Despite reporting inverted as being more difficult, subjects were equally accurate at this task for inverted and upright displays. Moreover, viewing inverted point-light walkers significantly increased gamma-band (25-30 Hz) MEG over the occipital areas. They conclude that the discrimination between upright, inverted and scrambled displays is “likely to be accomplished at relatively early stages, of cortical processing.” Evoked gamma enhancements reached a maximum at 100 ms after stimulus onset. Consistent with previous findings of an inversion effect, the behavioral data showed that subjects found the upright point-light displays more meaningful in the sense that the upright walker was rated as highly vivid, whereas the inverted displays received a significantly lower vividness rating. Pavlova et al. (2004) maintain that the early gamma band response registered by upright and inverted displays serves to dissociate the spatial coherence in these displays from the spatially scrambled displays. Later (130 ms and 170 ms) increases in gamma band response for the upright displays compared to the inverted displays were found in the right temporal lobe. Pavlova et al. (2004) assert that these increases that were obtained only for the upright displays reflected neural processing involved in gaining access to the meaningful structure when viewing those displays. So, it may be the case that while there is some level or degree of visual processing of coherent motion in inverted displays, there is no successful matching to stored dynamic templates. Early activation indicates this coherent form motion processing and later activation indicates access to meaning, i.e., identification, recognition, etc.

Jokisch, Daum, Suchan, and Troje (2005) used event related potentials (ERPs) to assess the level of cortical activation associated with viewing upright, inverted and scrambled biological motion. Subjects were given the task of pressing one key if they saw a point-light walker that was either upright or inverted and another key if the walker was scrambled. The results indicated that upright displays led to greater peak amplitude for the N170 component (peak amplitude within 150-200 ms latency window) than for the inverted displays. They interpret the activity associated with this early component as reflecting processing differences for upright and inverted displays, i.e., the inversion effect. In addition to this "early" component, Jokish et al. (2005) also found significantly greater amplitude in the N300 component for the upright and inverted walker than for the scrambled display. They suggest that the N300 reflects the top-down processing that is needed in order to potentially resolve visual ambiguity and illusory conjunctions. This kind of processing is thought to be associated with making more fine grained analyses between for example subordinate level objects or actions. Jokish et al. (2005) briefly discuss their results in the context of the results from Pavlova et al. (2004) and suggest that they are in accordance with one another. The occurrence of timing differences between the gamma activity and the ERP-components in the two studies is likely due to the different recording techniques and analysis methods.

Jokish et al. (2005) relate their findings to the global processing of upright displays and pop-out phenomena that require access to high-level cortical areas. They suggest that Hochstein and Ahissar's (2002) Reverse Hierarchy Theory (RHT) is a fitting theoretical framework from which to explain their data. The short latency of the N170 component is consistent with the fast feedforward processing associated with the pop-out effect. Much of the previously described evidence indicates that upright point-light displays are quickly processed as holistic dynamic gestalts, and subjects have early access to the categorical information, or meaning, of the stimuli. This what Hochstein and Ahissar (2002) call "vision at a glance." The later component, N300 is conjectured to reflect the top-down processing that is needed in order to potentially resolve visual ambiguity and illusory conjunctions. This kind of processing is thought to be associated with making more fine grained analyses between for example subordinate level objects or actions. Hochstein and Ahissar (2002) refer to this kind of processing as "vision with scrutiny." The relation between the visual processing of biological motion and the theoretical framework of RHT will be discussed further below.

6.5 Levels of Processing and Attention

The distinction between local and configural processing is of course related to the distinction between low- and high-level processing respectively. These distinctions also have their counterparts in terms of attention processes where active attention is associated with top-down conceptually driven processing and more passive attention is associated with bottom-up or stimulus driven attention (Thornton et al., 2002). Moreover, bottom-up, stimulus driven, attention occurs early on in visual processing whereas top-down, conceptually driven, processing occurs relatively late in visual processing.

A further aspect of the global processing of biological motion is the role that attention plays. The ability to see depicted actions in point-light displays of biological motion appears to require attention. Thornton et al. (2002) tested subjects using a dual task paradigm to assess the attentional demands in viewing a point-light walker. The primary task involved detecting the in-depth direction of the walker whereas the secondary task involved determining orientation changes for four rectangles that were displayed in the same dynamic mask as the point-light walker. The gist of their results revealed that performance on the secondary task was significantly reduced as the inter-stimulus interval (ISI) increased between the static frames of the sequence and as a function of type of dynamic mask (random or scrambled). The difference between a random mask and a scrambled mask is that in a random mask, the motion/trajectories of the individual masking elements and their spatial organization are random whereas in a scrambled mask, the motion of the individual masking elements matches the individual elements in the point-light target. So the difference between the motion pattern of the point-light target and the scrambled mask is that the scrambled mask is spatially scrambled, i.e., it does not share any spatial coherence with the target. When the point-light target walkers were separately masked by random and scrambled masks, there was a significantly greater reduction in performance associated with the secondary task when the walker was shown in a scrambled mask as compared to a random mask. Attention seems to play a role in processing displays of biological motion, and more attention is required when the displays are masked by scrambled elements.

It is important here to understand more specifically what Thornton et al. (2002) are claiming. They discuss the role of active and passive motion processing in the perception of biological motion. Passive motion processing is described in terms of low-level processing which takes place early in the visual pathway. This low-level processing is also more automatic in the sense that it places relatively less demand on attentional resources, as can be demonstrated by the intact performance of low-level processing when a secondary task is introduced. Active motion processing, on the other hand, is characterized as high-level. It is more susceptible to the attentional demands created by a secondary task. Active motion processing is also described as exhibiting a top-down level influence on low-level tasks or processes and requires more attention.

When it comes to viewing point-light displays without attention demanding secondary tasks, the processing is more passive, low-level and automatic. A shift from this level of processing to more active processing can occur if an attention demanding secondary task is introduced. The results from Thornton et al. (2002) show that "the human visual system can provide such efficient processing via at least two separate routes – a passive, automatic system that is affected only slightly by the withdrawal of attention (baseline, random-mask performance, experiment 1 and 2), and a top-down, active system that is much more dependent on the availability of attentional resources." (p. 851) I will return to this distinction shortly. In conclusion, there seems to be a wealth of evidence suggesting that the perception of biological motion in point-light displays is based on access to a global, holistic high-level representation of human motion. Results from Battelli, Cavanagh and Thornton (2003) showed that parietal patients with intact low-motion processing mechanisms had severe difficulties in a visual search task of biological motion displays. The obvious interpretation is that more high-level visual processing is necessary for biological motion perception. The high-level visual processing referred to here includes attention based integration of the different moving dots into a coherent, global percept, a kind of dynamic gestalt.

In a further series of experiments Thornton and Vuong (2004) extended their investigation of the respective roles of bottom-up and top-down effects and attentional processing. Using a flanker paradigm, they obtained results that showed an influence of passive bottom-up processing. Subjects were told to view a centrally located point-light walker and to report the direction in which the figure appeared to walk, either left or right. For some of the trials, however, the central target was surrounded, i.e., flanked, by 4 other point-light walkers that could either be walking to the left or to the right. This created situations where the direction of the flankers could be congruent or incongruent with the central target. The major issue was the extent to which the flankers would influence the time it took subjects to report the walking direction of the central target. It is important to mention that subjects were told to ignore the flanker stimuli. The gist of the results was that it took subjects significantly longer to report the walking direction of the central target when it was surrounded by incongruent flankers, i.e., point-light walkers walking in an opposite direction than when it was presented alone and when it was surrounded by congruent flankers. Subjects apparently incidentally processed the flankers to the extent that they had a negative influence on determining the walking direction of the central target.

According to Thornton and Vuong (2004), the incidental processing is achieved in a passive, bottom-up fashion. But instead of bottom-up processes being involved in local processing, the results suggested that bottom-up processing was involved in accessing the global motion of the flanker stimuli. This was demonstrated in another experiment where Thornton and Vuong (2004) created scrambled versions of the flankers, which had the effect of disrupting the global motion but preserved the local trajectories of the different walking directions. Under these conditions, there was no performance difference between congruent and incongruent conditions for the scrambled flankers, whereas for the "normal" flankers, the negative influence from the incongruent flankers was replicated. This shows that the bottom-up passive processing that leads to incidental processing of the flanker stimuli occurs when visual processing has access to global motion but not when access is limited to local motion processing. In this case, global motion processing can occur in the absence of more active and top-down controlled processing.

The gist of this research is that attention seems necessary for biological motion perception. The further question is what kind of attention is necessary. The previously mentioned results from Thornton and his colleagues indicate that at least passive attention is necessary. It is important to note, however, that attentional demands will be a function of the visual task. It should also be pointed out that the results from the experiments mentioned above have been limited to displays using an *upright* point-light walker. The role of attention in the processing of inverted displays of biological motion is largely unstudied.

Many of the previously mentioned findings on biological motion perception point towards *early* access to a global holistic *high-level* of processing, which is where semantic information is accessed. The semantic information in these cases has to with the ability to make category judgments between basic level categories. The suggestion that access is early and high-level seems somewhat contradictory if one takes the 'standard' starting point that low-level processes occur early and high-level processes occur late in visual perception.

Along with Jokish, et al. (2005), I suggest that Hochstein and Ahissar's (2002) Reverse Hierarchy Theory (RHT) could function as a useful framework from which to gain further insights about the levels of processing in biological motion perception. According to RHT, explicit perception is characterized by conscious access to recognition and identification. The reverse nature of the visual processing hierarchy is indicated by the idea that conscious visual perception *begins* at high cortical areas via initial feedforward mechanisms that implicitly follow a bottom-up hierarchical pathway. Top-down, or reverse hierarchy processing occurs *after* initial explicit perception and is characterized by the operations of feedback mechanisms in order to make fine grained perceptual discriminations like precise object/feature location, retinal size and color as well as component motion. The further claim of RHT is that explicit high-level perception is where basic level category judgments are made. Hochstein and Ahissar term this level of initial explicit perception as 'vision at a glance,' and it also reflects the activity of large receptive fields of high cortical areas and spread attention of initial perception. At the other (low-level) end of the processing continuum, 'vision with scrutiny' involves focused attention and the activation of small receptive fields in lower cortical areas. In contrast to previous ways of describing the temporal aspects of visual processing where high-level processing is deemed ‘late’ and low-level processing deemed ‘early’, Hochstein and Ahissar claim that *high-level processing* occurs *early* and *low-level processing* occurs *late*.

The idea here is that the visual quality of biological motion perception for upright displays is pop-out like, and this indicates global processing as well as quick access to semantic level representations. Consistent with Hochstein and Ahissar (2002), the perception of inverted displays could be characterized as an example of illusory conjunctions. There is some anecdotal evidence to suggest that people have difficulties in creating proper conjunctions of the individual points of light. So perception of inverted displays could be said to demonstrate the effects of top-down processing in the sense that the default value is an upright orientation and this creates false conjunctions in the perception of inverted displays. As Hochstein and Ahissar (2002) say,

> Thus, initial object recognition incorporates a priori “assumptions” influenced by experience. These features of initial high-level vision are a natural and direct out come of the receptive field properties of object-related neurons. (…) Thus vision with scrutiny is required to unbind initial incorrect conjunctions and revise vision at a glance when unexpected conjunctions are present tin the scene.” (p. 796)

As Hochstein and Ahissar (2002) indicate, “RHT (Reverse Hierarchy Theory) predicts that when attention is focused down to specific low-level cortical activity, default high-level detection may be compromised, and parallel activity may go unnoticed.” This line of reasoning is consistent with the reasoning in Shiffrar et al. (1997) where they show that global processes are involved in the perception of upright biological motion displays across apertures but that this global processing is impaired when inverted biological motion displays are viewed across apertures. Their findings show that global processing is associated with viewing upright displays and that local processing is associated with viewing inverted displays.

6.6 Summary and Further Empirical Issues to be Studied

The previously reviewed research suggests that action recognition is mediated by access to a cognitive representation of the recognized action. According to this view, a dynamic template of the recognized action serves as a reference from which to compare the visual input of observing another person performing an action. For example, Daprati, Wriessnegger and Lacquaniti (2007) assert that the observational *learning* of different actions relies on the construction and development of such a dynamic template. Orientation differences in the image plane make configural processing difficult. Access, however, to a dynamic action template is disrupted by image plane rotation, and differently oriented point-light walkers by 90 degrees or more are unable to prime one another (Pavlova & Sokolov, 2000). Not even prior knowledge about display inversion is sufficient to offset the negative effects of the different orientations (Pavlova & Sokolov, 2003). A central issue to be explored in the next chapter concerns the relationship between display orientation in the image plane and access to semantic information about the action category that is carried by a dynamic action template.

In addition to the previously mentioned findings that the visual processing of upright point-light displays includes access to semantic information or meaning, similar findings have been obtained by Boucart and Humphreys (1992) in the area of object identification. In a series of 7 experiments, Boucart and Humphreys (1992) investigated the relationship between visual global processing and access to semantic level information. The major issue was whether or not automatic access to semantic information occurs as a result of the global processing of familiar objects. The objects in their experiment were taken from two superordinate level categories, i.e., 6 vehicles and 6 animals. In their experiments, subjects were presented with a reference object, e.g., a sailboat. This reference object was presented as an outline drawing with a clear contour. The contour included information about characteristic parts of the object, e.g., wheels for some of the vehicles and legs for the animals. Shortly after viewing the reference object, a pair of fragmented objects was presented. Fragmentation consisted of breaks in the contour information. At this point, the subjects indicated which of the two fragmented objects matched the reference object according to the global shape of the objects. The important manipulation in terms of the present research was the semantic relatedness between the distractor in each pair and the reference object. Half of the distractors (which did not have the same global shape as the reference object) were semantically related to the reference object, i.e., came from the same superordinate category. The other half of the distractors were semantically unrelated to the reference object. In addition to this manipulation for familiar nameable objects, Boucart and Humphreys included nonnameable objects, which were distorted versions of the familiar objects. These nonnameable objects had the same global shape as their nameable versions, but the fragmented contour elements were rotated such that the colinearity of the elements was disturbed, which in turn led to severe naming difficulty. The logic of the experiments was as follows. If semantic level information is accessed in the task of matching for global shape, then subjects should take more time and make more errors when semantically related distractors are present than for semantically unrelated distractors when presented with nameable stimuli. For nonnameable stimuli, this difference should not occur because subjects are unable to access semantic level information about the identity of the presented object pairs. If subjects are unable to name the stimuli, then there should be no interference from semantic relatedness.

The results showed that subjects had automatic access to semantic information about object identification when they are required to attend to global physical shape defined by the orientation of the object's main axis. When subjects in the experiments were given the task of simply attending to the global shape of various objects, their matching performance was affected by the semantic relatedness of a simultaneously presented distractor. The effect of semantic relatedness was only found for nameable objects. When nonnameable objects were used, no effects of semantic relatedness were obtained. Boucart and Humphreys (1992, p. 804) conclude, "The results of the present series of experiments, however, show that semantic information can interfere with responses made on the basis of early visual codes." Hence, there is evidence in the area of object identification that global shape cannot be attended without object identification. In addition to the manipulation of "nameableness," Boucart and Humphreys included a manipulation that is critical to the experiment presented in the next chapter. As a further method of manipulating access to semantic information, they inverted the objects. Inverting the objects would disrupt semantic level processing and should therefore avoid any semantic interference when deciding which fragmented objects had the same global shape as the reference object. This is indeed what was found. It is important to note that Boucart and Humphreys assert that access to the global shape of the objects occurs when the objects are presented in an upright orientation and that this automatically leads to semantic level processing, i.e., information about object categories.

Given the inversion effect in the perception of biological motion processing, a similar line of reasoning can be used to investigate access to semantic (high-level) processing of point-light displays of different kinds of actions. If subjects view upright displays, then they should also have access to semantic/category level information about the different actions, i.e., the actions come from different action categories. When viewing inverted actions, on the other hand, previous research suggests that access to semantic level information is at least disrupted and therefore should prevent any ability to categorically distinguish between the actions depicted in the different displays.

Results from Grill-Spector and Kanwisher (2005) also support the notion that the ability to distinguish between basic level object categories occurs quickly and early on in visual processing. Subjects in their experiments viewed object images at 5 different exposure durations of 17, 33, 50, 68 and 167 ms. The images were masked immediately following presentation. Three separate tasks were to be carried out. For the object *detection* task, subjects were instructed to simply indicate whether or not a grey-scaled photograph contained an object. For the object *categorization* task, subjects were instructed to categorize the object in the picture at the basic level (e.g., car, house, and flower). The third task was also a categorization task (within category identification), but here subjects had to *identify* objects on a subordinate level of classification (e.g., kind of car). The purpose of the experiments was to see if the visual detection of objects precedes perceptual categorization. A further purpose was to investigate the extent to which objects are categorized on a coarser level (basic) before being identified "at a finer grain," e.g., subordinate level. If reaction time and accuracy differ for the different tasks, then it would seem that the different tasks require different processing times. Such a difference could also be interpreted as involving different mechanisms for the different tasks.

The results from Grill-Spector and Kanwisher (2005) revealed no significant differences between object detection and categorization (basic level). There were, however, significant differences between detection and categorization on the one hand and identification on the other. The identification task resulted in generally longer reaction times and lower accuracy when compared to detection and categorization. On the basis of the results, Grill-Spector and Kanwisher (p. 159) assert that "object detection and categorization performance are based on the same perceptual analyzers." Humans appear to have early access to semantic level information in the form of basic level object categorization.

In the research presented here, while not directly assessing the differences between active and passive processing, the active-passive distinction can be applied to the perception of upright and inverted displays of biological motion. Viewing upright oriented point-light displays of familiar actions under 'normal' viewing conditions (without increasing ISIs, or using apertures or dynamic masks) may only require passive processing, whereas viewing inverted displays will require active processing. So viewing upright displays will place less demand on attentional resources than viewing inverted displays. I think this way of interpreting the privileged processing of biological motion in terms of speed and accuracy demonstrated in previous studies best captures the data from a theoretical point of view. When people view upright point-light displays of biological motion, they are gaining quick (early) automatic access to high-level semantic representations of global motion patterns for human actions, perhaps as motion pattern neurons (Giese & Poggio, 2003) or as sprites (Cavanagh, et al., 2001). With regard to the role of attention, I am not taking the position that the perception of biological motion can be successful without recourse to attention. The view presented here, and in line with Reverse Hierarchy Theory, is that it is rather spread attention, not focal attention, that is needed to perceive biological motion under standard or normal viewing conditions.

The claim here is that global, high-level processing in the perception of biological motion is characterized by (phenomenally) direct access to the categorical nature of the motion presented in a point-light display. The categorical nature of the display is the basic level action depicted in the display. The next chapter describes an experiment that directly tests the relationship between access to high-level semantic information about action category and action orientation.

Chapter 7 - Automatic Activation of Category Information and the Inversion effect in Displays of Biological Motion[41]

Behavioral and neuroscientific results from experiments on the visual processing of upright and inverted displays suggests differential access to stored high-level representations and/or different processing mechanisms that mediate perception of upright and inverted displays of biological motion. More specifically, findings from experiments on biological motion perception indicate the following differences in the visual processing of upright and inverted displays. For upright (non-masked) displays, visual processing:

- is fast and "automatic" (indicates pop-out) (Jokisch et al., 2005; Giese & Poggio, 2003);
- involves high-level global processing mechanisms (Bertenthal & Pinto, 1994; Shiffrar et al., 1997);
- involves access to categorical information (Dittrich, 1993, Pinto & Shiffrar, 1999) and
- requires attention (Battelli et al., 2003; Thornton et al., 2002; Hirai et al., 2005).

There is an apparent conflict between the first and last points in this list. While I acknowledge the apparent conflict, I am not prepared to state that they are necessarily mutually exclusive. The conflict might be more apparent than real. Recall that Hochstein and Ahissar (2002) describe vision at a glance as utilizing spread attention to capture the gist of scene or to detect objects at a basic level of description. Spread

---

[41] The experiment presented in this chapter was carried out by Sigríður Pálsdóttir who did her senior thesis on the topics presented here. Many thanks to Sigríður for her work. The contents of this chapter have been presented previously at 3 different conferences (Hemeren, 2003, 2005; Hemeren & Pálsdóttir, 2003): 26th European Conference on Visual Perception, Paris 2003; XII Conference of the European Society of Cognitive Psychology, Granada, Spain 2003 and XXVII Annual Meeting of the Cognitive Science Society, Stresa, Italy 2005.

attention is thought to initially guide the initial feedforward processing in visual perception. In this sense, biological motion perception can be both automatic and require attention. I will discuss this issue further in Chapter 7.

Concerning the relationship between global processing mechanisms and access to categorical information, results from object recognition studies indicate that access to the global shape of *static* objects automatically activates identification (Boucart & Humphreys, 1992). Therefore, to the extent that displays of biological motion represent *dynamic* objects, information about the categorical nature of the depicted actions may be automatically accessed if visual processing occurs on a global level.

In relation to the factors characterizing the processing of upright actions and for the purpose of the work presented here, there is evidence to suggest that the visual processing of inverted actions:

- is slower and indicates less (if any) pop-out (Dittrich, 1993; Pavlova & Sokolov, 2000);
- impairs accurate high-level global processing and appears to rely more on local motion processing (Pavlova & Sokolov, 2003; Pinto & Shiffrar, 1999) and
- impairs access to categorical level information (Pinto & Shiffrar, 1999).
- There is no specific data on the role of attention in the visual processing of inverted displays.

The central issue to be investigated in the experiment in this chapter concerns the relationship between display orientation and visual access to categorical information associated with different dynamic action templates. More specifically, if categorical information is automatically activated, then effects of that information should be seen to a greater extent for upright displays than for inverted displays. This follows from the previously mentioned findings on the inversion effect, which disrupts access to configural/global information. In order to assess the extent to which categorical information is used by subjects, different kinds of actions will have to be used as stimuli. Much of the biological motion research has used the point-light walker as the primary stimulus. Relatively few studies have systematically investigated the potential differences between different actions, or action categories, and the extent to which effects of orientation specificity may vary depending on action category. Recall that findings from Dittrich (1993) show that subjects' ability to identify different point-light actions was differently affected by inversion. For example, identifying the actions of *boxing* and *dancing* was easier than identifying the actions of *greeting* and *threatening* when they were presented upright. However, identification was easiest (although diminished) for *dancing* and *greeting* when they were inverted, which suggests that the inversion effect interacts with action category. Three patch-light actions (walking, climbing rope and jumping jacks) will be used in this experiment. The choice of these actions will be further discussed in the materials section.

The relationship between display orientation and access to categorical information will be investigated by determining the extent to which the visual processing of a patch-light action might facilitate, i.e., prime, the later visual processing of the same or other patch-light actions. Pavlova and Sokolov (2000) used a priming paradigm where subjects viewed the unmasked prime for 10 seconds, and subjects were also informed about the relatively different orientations for the priming and primed displays. Recall that previous results from Pavlova and Sokolov (2000) showed that prior exposure to an *unmasked upright* point-light walker led to an increase in detection performance only for *masked upright* point-light walkers. There was no facilitation in detection performance when an *upright* display preceded an *inverted* display. Prior exposure to an unmasked *inverted* display had no effect on later detection of a masked *upright* display, and had no effect on later detection of a masked *inverted* display.

The lack of a priming effect for incongruent display orientations, i.e., upright-inverted, inverted-upright, indicates for example that the visual processing of upright oriented point-light displays does not facilitate the visual processing of inverted displays. It could be the case that this priming paradigm and the task of detecting a point-light target within a dynamic mask are not conducive to obtaining a priming effect with incongruent primes and primed point-light actions. Given a different method and task, it may be possible to obtain some priming of incongruent displays.
There is, however, some evidence that suggests that priming effects might be obtained with inverted displays, especially when the priming and primed displays are both inverted, i.e., orientation congruent. Firstly the inversion effect is about the *relative* processing differences between upright and inverted point-light displays. Although the visual processing of inverted displays is impaired relative to upright displays, there appears to still be some level of processing that occurs when subjects see inverted displays (Grossman & Blake, 2001; Pavlova et al., 2004) and when subjects learn to detect the presence of inverted point-light targets (Grossman et al., 2005; Hiris et al., 2005). Can this level of activation and processing for inverted displays lead to significant priming when inverted displays are presented as primes?

In order to test for the effects of categorical information and display orientation on the visual processing of point-light displays, subjects will be exposed to repetition priming. This will be in contrast to the long-term priming paradigm that Pavlova and Sokolov (2000) used. Effects of viewing different point-light actions upright and inverted on multiple occasions may lead to facilitation for the point-light displays that immediately follow the priming displays, even when the primed displays are inverted. Despite the numerous articles on the inversion effect for point-light displays, there seems to be no previous experiments that have used a repetition priming paradigm. Verfaillie (1993, 2000), however, used repetition priming to investigate orientation dependent processing for displays rotated in depth. In this case, he used left and right facing point-light walkers. A further manipulation included differences in direction of articulation for the walkers, i.e., either forward or backward walking. The results showed that congruently oriented in-depth displays (both left facing or both right facing) led to significantly more priming that than incongruent displays, which supports previous findings of orientation specificity. However, there was still an observable priming effect when right facing walkers preceded left facing walking and vice versa (Verfaillie, 1993 exp. 6; 2000). A finding of significant recognition performance for incongruently oriented point-light walkers was replicated by Troje, Westhoff and Lavrov (2005). Given the proven effectiveness of the repetition priming paradigm in Verfaillie (1993) I chose to use it as the methodological basis for the experiment in this chapter and will refer to Verfaillie's original work periodically. The effectiveness of the paradigm was shown by its sensitivity. Only 6 subjects were needed to obtain statistically reliable priming effects.

As for many experiments, the choice of task is critical to the interpretation of the results. What are subjects being asked to do, and what kinds of cognitive processes are supposedly needed to adequately perform the task? In this experiment, an orientation decision task (Boucart et al., 2000) will be used. The gist of this task is to indicate whether the patch-light action sequence is upright or inverted. A crucial question here concerns the extent to which the task requires access to semantic level (categorical) information and whether or not that access is automatic, i.e., access occurs without conscious effort to obtain that information. Boucart et al. (2000) showed that an orientation decision task for objects automatically activated semantic level processing as revealed by significant activation of cortical area 37 (Brodmann), which according to Boucart et al. has previous been found to be critically involved in object identification.

In order to successfully perform the orientation decision task, visual processing requires at least the detection of some local (motion or form based) configuration of patches. The next step would then be to determine the orientation of the local configuration. For example, the local configuration of patches for the leg or arm of a human figure might be detected and then depending on its relation to the rest of the patches be judged to be either upright or inverted. In this case, I want to leave open the possibility that determining the orientation of a display can be done without access to global/configural information, and therefore without access to information about high level knowledge of the action category. This explanation would be consistent with the explanation and findings from Troje and Westhoff (2006) about obtaining the inversion effect on the basis local motion patterns of the feet.

It may, however, also be the case that the orientation decision task relies on access to configural processing of the whole object, which in this case is a human body. Since previous research has shown that configural processing is disrupted for inverted displays, subjects may have difficulty determining the orientation of the displays.

It seems reasonable to assert that a correct decision about the orientation of an inverted display requires access to some identifying information about the human body. This is supported by the previous results from Pinto and Shiffrar (1999) who found that subjects were able to identify inverted displays as depicting a human body, although subjects identified significantly more upright displays than inverted displays. It was also the case that when displays only consisted of different subconfigurations of the human body, detection did not differ from viewing upright displays. This suggests that even subconfigurations of the human body function as reliable indicators of the global motion of the human body and actions performed by it.

By using different patch-light actions and presenting them upright and inverted, it will possible to assess the extent to which the orientation decision task leads to the automatic activation of categorical information. If subjects are faster at determining the orientation of an upright walking action when it is preceded by an upright walking action compared to an upright climbing action, then it would appear that subjects are making categorical distinctions between an action that primes itself and an action that is primed by another action. I should point out that I am not claiming that access to categorical information is explicit. It is rather incidental or implicit since there is no explicit recognition or identification procedure. It seems, however, reasonable to suggest that automatic implicit access to categorical information would be indicated if a specific action is better at priming itself than it is at priming other actions.

The occurrence of the inversion effect in this experiment should be seen in results that show that automatic implicit access to categorical information for the upright displays mentioned above will be disrupted for inverted displays. In contrast to upright displays, an inverted point-light walker will be no better at priming itself than it will be at priming other actions, if in fact there is any significant level of priming at all. In general, orientation congruent priming and primed displays should lead to greater levels of priming than orientation incongruent priming and primed displays. According to previous results from Pavlova and Sokolov (2000), there should be no facilitation for the orientation decision task when the priming and primed displays have different orientations. The following methodological details specify more clearly how the experiment was constructed and carried out.

7.1 Method

7.1.1 Subjects

Eight students (4 females and 4 males) from the University of Skövde participated in the experiment. (Age: range 21-24 yrs.) All subjects had normal or corrected-to-normal vision. One subject was familiar with the nature of the experiment whereas the other subjects were naïve regarding the nature of the stimuli and that priming effects would be investigated. Seven subjects said they were right-handed. Subjects received two tickets to the movie theatre as compensation for their participation. Subject participation conformed to the ethical guidelines established by the Swedish Research Council (2002).[42]

7.1.2 Materials

7.1.2.1 Apparatus

The stimuli were displayed on a Macintosh 17” (33 x 25 cm) monitor set to black and white with a resolution of 832 x 624 pixels and a refresh rate of 75Hz. Stimulus presentation was controlled by a Macintosh 7100/66AV (66 MHz). A standard Macintosh keyboard was used to register subject response. DotPlayer recorded subject response and reaction time (ms) with a ± 4ms margin of error.

7.1.2.2 Stimuli

Patch-light Actions

The basic technique for recording and manipulating the patch-light stimuli used in this experiment was described in Chapter 4. This section will describe the relevant details needed to understand how the stimuli were used in this specific experiment.

In addition to walking, two additional actions were included in the experiment (Fig. 7.1). Climbing a rope and jumping jacks were included to investigate the generalizability of previous results using only a point-light walker. The previous results referred to here concern the inversion effect. It might be the case that the inversion effect is a matter of degree which varies depending on the kind (category) of action shown. A point-light walker, for example, exhibits dynamic symmetry, and other actions are not quite so dynamically symmetrical. Throwing, climbing up a rope and waving are actions that are not as obviously dynamically symmetrical as walking or doing jumping jacks. Climbing up a rope also differs in regard to the surface supporting the action. Whereas a hard surface supports walking, a rope is used to support climbing. In addition to the 3 actions, 2 neutral stimuli were created to establish a neutral baseline. The creation of the neutral stimuli will be explained in the next section.

The translation components in the walking and climbing displays were removed. For the jumping jacks, there was no translational component. The figure performed the jumping jacks without moving across the floor. For the patch-light walker, a maximum of 10 patches were visible during the sequence (Figure 7.1). These visible patches were attached to the head, the right shoulder, elbow and hip, both wrists, both knees, and both ankles. While the patches on the elbow and the left shoulder were always occluded, the patches on the left ankle, knee and wrist were temporarily

---

[42] The ethical guidelines are in Swedish, and the reference will therefore be indexed according to the Swedish name for the Swedish research council, namely, Vetenskapsrådet. See the references for the full reference.

occluded during appropriate phases in the step cycle. The direction of articulation was to the right of the viewer. The visual angles for the height and width of the patch-light walker varied according to the vertical and horizontal extension of the body during the action. The visual angle for the height ranged from 4.03° to 4.59°, and the visual angle for the width ranged from 1.08° to 2.17°.

A whole action cycle of two steps was completed in 26 frames, which at the speed of 648 ms per step resulted in a natural looking version of a walking person. At this speed, the walker would complete 46 cycles per minute, which is in accordance with the normal walking speed of 30 to 70 cycles per minute (Imman, Ralson & Todd, 1981, cited in Pavlova & Sokolov, 2000). Of the three actions, it took the walker the shortest time to complete an action cycle, the duration of the other actions was based on that time and all actions were shown for 26 frames. The frame display rate was set to 20 frames-per-second, which resulted in a display duration of 1.3 seconds for each action.

For the sequence of climbing up a rope, there were no patches that were occluded throughout the whole sequence (Figure 7.1). The patch that marked the hip and the left shoulder were only visible for three frames in the sequence. The orientation of the patch-light climber differed somewhat from the patch-light walker. Though the patch-light walker faced the right at about 90°, the figure in the climbing sequence had more of a three-quarter view towards the right, about 45°. The reason for this was that this orientation was thought to be optimal in terms of being able to reduce occlusion and yet maintain access to velocity information along the horizontal and vertical dimensions. The problem of deciding what the best orientation for each action is difficult to solve without more systematic investigation of the issue. The pragmatic solution that was used in this experiment was to simply choose the orientation that seemed most perceptually advantageous for each specific action, the consequence of which is that the three different actions will be presented from three somewhat different orientations. The visual size in terms of the viewing angle of the patch-light display for climbing up a rope also varied in height and width. The height of the figure varied between 3.61° and 5.18°, and the width varied between 1.25° and 1.74° of viewing angle.

The action of doing jumping jacks was not rotated in depth. The jumping jack figure was shown facing completely to the front, and 12 patches were fully visible during the sequence (Figure 7.1). Given the motion of the wrists and feet, as well as their movement in relation to the major axis of elongation of the human body, the fully frontal perspective was deemed to provide the view with the most information regarding the relative velocities of the limbs and their relative movement to one another. The height and width of the jumping jacks figure varied also during the sequence. The height of the figure varied between 3.90° and 4.69°, and the width varied between 1.25° and 4.48° of viewing angle.

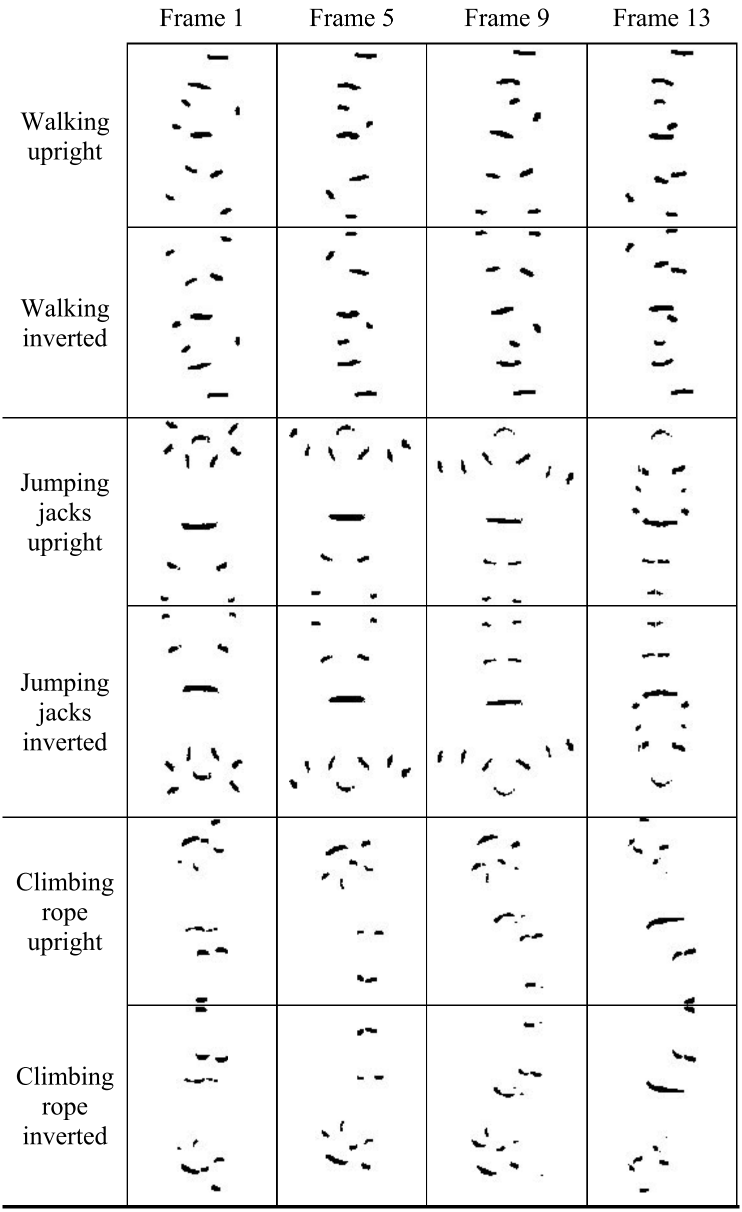


Figure 7.1. Frames from the three patch-light actions, upright and inverted, that were used as stimuli in the experiment.

All actions were also inverted such that they had the same direction of articulation. If the upright displays are simply rotated 180 degrees, the spatial relations are left-right and up-down reversed relative to the upright view. Only the positions of axes of symmetry and the main axis of elongation remain the same. By flipping the display in the sagittal plane, left-right spatial relations are held constant between upright and inverted displays. This also has the effect of holding the direction of articulation constant. All other variables such as size and display duration were held constant across the upright and inverted displays.

A further step was taken to prevent the recognition of the different actions by simply discovering a unique pattern on the first frame of each sequence. Three different starting points for each upright action sequence were selected. The starting points differed by 5 frames from one another. It should be mentioned, however, that the inverted displays had the same starting points as the upright displays.

Neutral (baseline) Stimuli

Two neutral stimuli were constructed as the basis from which to measure any potential facilitation of priming stimuli. The reason for creating 2 neutral stimuli was to create the same number of response alternatives for the neutral stimuli as for the patch-light actions (upright or inverted). In this case, subjects will have to indicate which of the 2 neutral stimuli are presented on a given trial. The creation of the neutral stimuli in this experiment was based on Verfaillie's (1993, 2000) description of the methods he used to create neutral stimuli for use in his repetition priming studies.

The neutral stimuli were composed of 7 patches taken from the three different patch-light actions (Figure 7.2). The vertical midpoint of both neutral stimuli consisted of 3 horizontally lined up patches, which remained stationary throughout the motion sequence. Two of the remaining 4 patches were vertically positioned above the midpoint, and the other 2 patches were vertically placed below the midpoint. The motion components of the neutral stimuli were determined by the motion of the 2 pairs of patches above and below the midpoint. The *local* motion trajectories of these pairs of patches were the same for both neutral stimuli. The *global* motion, however, differed. For both stimuli, the patch directly above the center patch in the midpoint rotated back and forth between 50° and 310° (with 0° at 12 o'clock), and the patch directly below the center patch rotated back and forth between 130° and 230°. The remaining two patches at the top and bottom of the stimuli moved back and forth in a straight pathway. A further aspect of the local motion trajectories was the local motion of the patches above and below the midpoint. The 2 patches in each pair moved in the same direction *relative to one another*, but as a pair, they did not always move in the same direction *relative to the other pair*. This difference reflects the need for global processing because visual processing of both pairs above and below the midpoint is needed to distinguish the relative motion path differences between the patch pairs. For neutral stimulus N1, the patch pairs moved in opposite directions, whereas for neutral stimulus N2, the patch pairs moved in the same direction (Figure 7.2).

Another way of describing the difference between N1 and N2 is to say that N2 was created by simply creating two top patch pairs from each frame in N1and then flipping one of the pairs and placing it under the horizontal midpoint in each of the frames. The end product of this procedure is a pair of neutral stimuli that require global processing and move in a partially rigid way. The height and width of the 2 neutral stimuli was determined by calculating the mean for the height and width of the other patch-light displays and then determining the visual angle based on a viewing distance of 70 cm. The visual angle for the height of the neutral stimuli was 4.33° and the width was 1.97°. The height and width remained constant during the motion sequence. The neutral stimuli completed one and a half cycles during the 26 frame duration. A cycle was defined as one complete oscillation from left to right and back again. As for the patch-light actions, 3 different files with different starting points for the neutral stimuli were also created to prevent the case where task performance is merely a function of visual processing of the first image in each sequence. The 3 different starting points for N1 and N2 were the same.

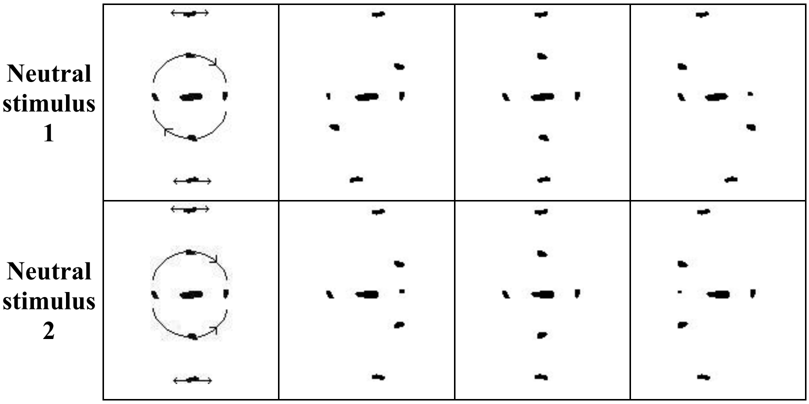


Figure 7.2. Frames from the sequence of the neutral stimuli. The arrows in the first frames show the movement of the patches during the sequence. The three following frames show the full extent of motion of the patches for each neutral stimulus.

7.1.3 Design and Procedure

The central experimental feature of the design of this priming experiment is the creation of transitions that consist of priming and primed patch-light action sequences. The reaction time in milliseconds was measured for each transition. The gist of the experimental design then is the creation of the transitions that will reflect the issues being studied. To this end, the design of the experiment includes 4 independent variables. Two of these variables and their combinations have to do with the investigation of the inversion effect. These two variables are *orientation congruence* and *prime orientation*. Orientation congruence consists of two levels: congruent and incongruent. The priming and primed stimuli are either orientation congruent or orientation incongruent. Prime orientation consists also of two levels: upright and inverted. The priming stimulus is either upright or inverted. The crossing of these two variables leads to the following combinations of priming and primed displays:

- **upright – upright**: orientation congruent and priming display is upright,
- **inverted – inverted**: orientation congruent and priming display is inverted,
- **upright – inverted**: orientation incongruent and priming display is upright,
- **inverted – upright**: orientation incongruent and priming display is inverted.

The other 2 variables are *primed action*, which has three levels, i.e., walking, jumping rope and climbing up a rope, and the *action congruence variable*, i.e., the same action either occurs as the priming and primed action or the action is primed by one of 2 remaining actions. Consequently, the action congruence variable has 3 levels: congruent, incongruent and incongruent. There are two levels of incongruence because an action can only be congruent with itself, but it can be incongruent with the other two actions. Table 7.1 shows the combinations that constitute the conditions that are constructed by crossing the two variables.

When the variables in Table 7.1 are crossed with orientation congruence and prime congruence, the result is a 2x2x3x3 design where all variables are manipulated within groups. Therefore, there will be a total of 36 transition conditions in the experiment.

Table 7.1. Combinations of the levels for the independent variables of action and action congruence.

	Primed Action		
Action congruence	*climbing rope*	*jumping jacks*	*walking*
congruent	climbing rope	jumping jacks	walking
incongruent	jumping jacks	climbing rope	jumping jacks
incongruent	walking	walking	climbing rope

Note: Each of the 9 cells represents a condition for the kinds of priming - primed transitions in the experiment.

Subjects participated individually in 5 sessions. The sessions were distributed over a period of 6 days, and no subject participated in more than one session per day. Prior to the first session, subjects were informed about the general nature and procedure of the experiment. Subjects were told that they would view 5 different patch-light sequences in random order and that 3 of the sequences represented actions performed by a human actor (climbing up a rope, a person doing jumping jacks and walking). Each action would be presented in an upright orientation and inverted. In addition to the 3 human actions, 2 abstract, or neutral, patch-light sequences would also be presented. These sequences would be presented many times throughout the five sessions. When presented with a patch-light sequence, subjects were instructed to simply indicate whether they thought the sequence was upright or inverted, i.e., a serial two-choice reaction time task. Following the instructions, subjects completed a practice session to familiarize themselves with the task of making the correct key presses. The stimuli in the practice session were the same as the stimuli in the experiment. The experimenter was present during the practice session to correct any misunderstandings and to answer questions. Both speed and accuracy were emphasized to the subjects. During the experiment no response feedback was given to the subjects.

Subjects were seated in a dimly lit room with a viewing distance of 70cm to the computer screen. They were also told to maintain the viewing distance throughout the different sessions. A measuring tape was provided so that the subjects could check the viewing distance at the start of each session and sub session. The stimuli were presented in the center of the computer screen and subjects were informed that the stimuli would always be presented there. Following the subject response, a response-stimulus interval (RSI) of 500 ms occurred, during which the display was white, i.e., the same color as the background for the patch-light stimuli.

Each of the 5 sessions was divided into two sub-sessions, and each sub session contained 6 blocks of trials. One block contained 144 trials where each action, orientation and neutral display occurred 18 times. Each sub session therefore contained 864 trials, and each session contained 1728 trials. A total of 8640 trials were completed by each subject after the 5 sessions. Trials within blocks and block order within each sub session were randomized for each subject.

Stimulus configurations appeared one at a time in a random order. Each motion was viewable for up to a maximum of 1300 ms, after which followed a blank screen. Subjects were to respond to each action sequence by determining whether the sequence was upright or inverted. For the neutral stimuli, subjects were instructed to distinguish between them by indicating whether they saw N1 or N2. A subject response that occurred before the end of a motion sequence terminated the sequence and started the RSI. Responses were indicated by pressing either the left-arrow key or the right-arrow key on the key board. Left-right key presses were counter-balanced across subjects. Half of the subjects (2 males and 2 females) pressed the right arrow key for upright stimuli and the left arrow key for inverted stimuli. This order was reversed for the other 4 subjects. The key presses for the neutral stimuli were counter-balanced within each of these groups. So, subjects that pressed the right arrow key for upright stimuli were divided into two groups where one group (one male and one female) pressed the same key for the neutral stimulus N1, whereas the other group pressed the right arrow key for the neutral stimulus N2. This balancing was also carried out for the other half of the subjects. This counter balancing resulted in subjects pressing the right arrow key as many times as the left arrow keys for correct responses.

7.2 Results

The results are presented in two parts. In the first part, an analysis of the reaction times (RT) as a function of display orientation and type of action (walking, jumping rope and jumping jacks) independent of potential priming will be presented. The purpose of this analysis is to see if there is a difference between the upright and inverted displays, which is expected as an instance of the inversion effect. The other purpose is to determine the extent to which RT differences occur for the different actions. Of particular interest is the potential interaction between display orientation and type of action. One of the reasons for the experiment was to gather data about the extent to which orientation specific effects might depend on the action category used in the experiment. The second part of the results section will analyze the potential priming effects obtained in the experiment and will also include a description of how the neutral stimuli were used to calculate the neutral baseline.

Since this experiment follows the methods presented in Verfaillie's (1993, 2000) previous experiments, the initial treatment/sorting of the data will also follow suit with Verfaillie. For each participant, data from the first session and the first two blocks from each subsequent session as well as the first five trials in each remaining block served as training and were excluded from any analyses. Incorrect responses from the remaining 34,624 trials were eliminated. Mean RT and standard deviation were then calculated for each participant. In addition to the incorrect responses, RTs exceeding the mean by three standard deviations were then excluded, leaving 93% of the data as the basis from which the following results have been determined.

7.2.1 Analysis of Reaction Times

The mean RTs in milliseconds for correct answers and for each action and orientation are presented in Table 7.1. The RTs show that subjects responded fairly quickly. The display duration of 1300 ms was more than enough to make a reliable decision about the orientation of the displays. A repeated measures analysis of variance (ANOVA) with display orientation (upright and inverted) and action (climbing rope, jumping jacks and walking) as the independent variables was performed on the data.

A look at the means indicates that subjects took more time to make the orientation decision for inverted displays (M = 436 ms) compared to upright displays (M = 409 ms). This main effect of orientation was statistically significant, $F(1,7) = 16.67$, MSE

= 525, partial $\eta^2 = .708$, $p = .005$. It also appears that main effect of action shows some differences between the RTs for the different actions. This main effect was also statistically significant, $F(2,14) = 24.21$, MSE = 76, partial $\eta^2 = .776$, $p < .0001$. Post-hoc Bonferroni adjusted comparisons for the main effect of Action showed that subjects responded significantly faster to jumping jacks (M = 411 ms) than to both climbing rope (M = 431), $t(7) = 6.65$, $p = .001$, and walking (M = 427), $t(7) = 4.21$, $p = .012$. The difference between climbing rope and walking was not significant, $t(7) = 1.71$, $p = .40$. The interaction between orientation and action was not statistically significant, $F(2,14) = 1.08$, MSE = 22, $p = .37$. This lack of an interaction effect shows that the effect of orientation did not vary as a function of the different actions. All three action sequences (in this experiment) seem to be effected equally by differences in orientation. Performance decreased to a similar extent for all actions when they were presented in an inverted orientation.

Table 7.2. Mean reaction times in milliseconds to make an orientation decision as a function of action and orientation. Standard errors of mean are in parentheses.

		Action		
		climbing rope	jumping jacks	walking
Orientation	upright	418 (11)	398 (13)	413 (11)
	inverted	443 (12)	424 (13)	442 (11)

Even given the short reaction times of the data, around half a second, and the "simple" task of making an orientation decision, subjects appear to at least be able to discriminate between jumping jacks and the other two actions, making a rudimentary categorical discrimination. The results also confirm what many other studies have shown, namely, an effect of orientation. Even though the difference in RT was significant, it was not large: 27 ms. Perhaps it is not so surprising that this difference in RT is so small. The conjecture here is that the behavioral task does not demand that subjects make discriminations between actions, a semantic or categorization task. The only thing subjects had to do was to be able to determine whether or not a presented display was upright or inverted, an orientation decision task.[43]

---

[43] Verfaillie (1993) used an object-decision task where his subjects made a decision about whether the display depicted a human or non-human walker and claimed that subjects did not need to interpret local features or relative motion of the body parts and thereby avoided processing information about the direction of articulation.

7.2.2 Assessing Repetition Priming Effects

Priming in this experiment is defined as the effect of a stimulus, n-1, on the RT (orientation decision) of a directly following stimulus, n. Previous exposure to a stimulus facilitates or inhibits later processing of same or similar stimuli. Priming effects in this experiment were assessed by determining the effects of stimuli that followed one another in quick succession. Figure 7.3 presents the basic steps for calculating the priming effects in this experiment.

- Neutral priming stimulus → primed stimulus_1 (reaction time_1 neutral baseline)
- $\text{Priming stimulus}_1$ → primed stimulus_1 (reaction time_2: same action and orientation are used as priming and primed stimulus)
- $\text{Priming stimulus}_2$ → primed stimulus_1 (reaction time_3: different action and/or orientation are used as priming and primed stimulus)
- reaction time_1 - reaction time_2 = priming effect of stimulus_1
- reaction time_1 - reaction time_3 = priming effect of stimulus_2

Figure 7.3. Basic steps for calculating the priming effects.

Each action (climbing rope, jumping jacks and walking) served as both priming and primed stimulus for each of the other actions, resulting in 9 different action transitions (action congruence transitions). There were also four orientation transitions for the priming and primed stimuli: upright-upright, inverted-inverted, upright-inverted and inverted-upright. The combination of these transition types results in 36 total transitions.

7.2.2.1 Calculation of Baselines

In order to assess possible effects of priming, a baseline needed to be established. As mentioned previously, neutral priming stimuli were included in the experiment for just this purpose. In order to obtain a sensitive assessment of priming effects, a baseline should take into consideration possible differences between effects of the different actions, effects of orientation, and differences in how subjects are instructed to respond. Ideally, the same baseline could be used to examine the priming effects associated with all conditions. It is rarely the case, however, that such an ideal condition exists. The time it takes to make an orientation decision for the different patch-light actions when they are preceded by the neutral stimuli would be likely affected by previous experience and the kinds of responses that need to be made, i.e., key-press sequence.

The basic logic of calculating the baselines is based on the obtained RT to make an orientation decision for an action when it is immediately preceded by one of the two neutral primes. However, in order to obtain an appropriately sensitive baseline, it may be necessary to take into account the extent to which possible differences between the different actions, orientations and key-press sequences affect RT *prior* to any priming analyses. If, for example, an analysis reveals that subjects respond differently (faster or slower) due to the different key-press sequences, then separate baselines should be calculated based on this difference. The same reasoning applies to the different orientations and actions. For example, we would expect the need for at least two different baselines that reflect the differences between reacting to upright vs. inverted stimuli. It would be unfeasible to establish a baseline for upright oriented stimuli on the basis of responses to inverted stimuli. Likewise, subjects may respond differently to climbing a rope, jumping jacks and walking as primed stimuli even though they are preceded by the same neutral stimuli. If this is the case, then the baseline for the orientation decision for jumping jacks will consist of only those trials where jumping jacks was preceded by the neutral stimulus. Consequently, the baseline for evaluating the priming effect when viewing an inverted patch-light walker will be the average time it takes a given subject to decide whether an inverted walker is inverted *and* when it is immediately preceded by a neutral stimulus *and* when the sequence of key-presses is the same.

One aspect to keep in mind when discussing the baselines is that there is no bias or methodological problem with calculating separate baselines for each factor even if there are no differences due to the above mentioned factors. The only disadvantage is that it would be unnecessary. Whereas, failing to calculate separate baselines when there are in fact differences would result in serious methodological problems in assessing appropriate priming effects. In this case, priming, or the lack of it, could/would quite likely be due to the fact that the baseline overestimates or underestimates an eventual priming effect due to differences in the conditions under which the priming and primed responses occur.

The mean reaction times to make an orientation decision when the actions are immediately preceded by one of the neutral stimuli are presented in Table 7.3.

Table 7.3. Mean reaction time in milliseconds for baseline stimuli as a function of Key Press, Action and Orientation. Standard error in parentheses.

	Same Key Press				Different Key Press		
	Action				Action		
Orientation	climb rope	jumping jacks	walking	Orientation	climb rope	jumping jacks	walking
upright	438 (11)	424 (11)	433 (11)	upright	431 (12)	407 (13)	430 (11)
inverted	488 (15)	457 (16)	469 (12)	inverted	462 (10)	443 (13)	460 (12)

A 2 (key-press: same-different) x 2 (orientation: upright-inverted) x 3 (action type: climb rope, jumping jacks and walking) repeated-measures ANOVA (univariate) was carried out to examine the extent to which separate baselines need to be calculated for the different orientations, actions and associated key press sequences. The analysis revealed significant main effects of key-press sequence ($F(1,7) = 6.69$, MSE = 553, $\eta^2 = .489$, $p = .036$), orientation ($F(1,7) = 14.91$, MSE = 2096, $\eta^2 = .680$, $p = .006$), and action type ($F(2,14) = 11.78$, MSE = 342, $\eta^2 = .627$, $p = .001$). Subjects responded more slowly when key presses were the same for the neutral prime and primed action than when they were different (451 ms vs. 439). Orientation decisions for upright primed actions were faster than inverted (427 ms vs. 463). Means for the main effect of action were 455 ms for climbing a rope, 433 ms for jumping jacks and 448 ms for walking. No further a posteriori analyses were performed for the action types because the purpose of the purpose of the analysis is simply to establish the potential difference between the different actions. The interaction between key-press sequence and action was also significant ($F(2,14) = 5.72$, MSE = 47, $\eta^2 = .450$, $p = .015$). No other interactions were significant.

Since all three main effects were significant, 12 neutral baselines reflecting these effects were calculated for each participant. For each action 4 separate baselines were calculated; 1) when key presses for neutral prime and primed actions were the same and the primed actions were inverted, 2) same key-presses for neutral prime and primed actions but the primed actions were upright, 3) key-presses were different for neutral prime and primed actions and primed actions were inverted and 4) key-presses were different for neutral prime and primed actions and primed actions were upright. These baselines were then used to assess the priming effects presented below. More specifically for example, the neutral baseline for evaluating the priming of an upright display of climbing a rope was the mean (for a specific subject) orientation decision RT for an upright display of climbing a rope *when* it was immediately preceded by a neutral stimulus. It also had to be the case that the sequence of key-presses was the same for the neutral baseline and the priming-primed transition.

7.2.2.2 Priming Analyses[44]

The mean priming effect in milliseconds for each of the 36 transitions mentioned in section 7.1.3 were calculated on the basis of the steps mentioned in the previous section. The results are presented in Figure 7.4. Firstly, there are three clearly evident trends in the data. The first is that there appears to be more priming for orientation congruent transitions than for orientation incongruent transitions. This can be seen by comparing the height of the bars for Panels A and B with the height of the bars (despite the different scale) for Panels C and D.

---

[44] Many thanks to Júlía Pálmadóttir for programming the macros that sorted the data.

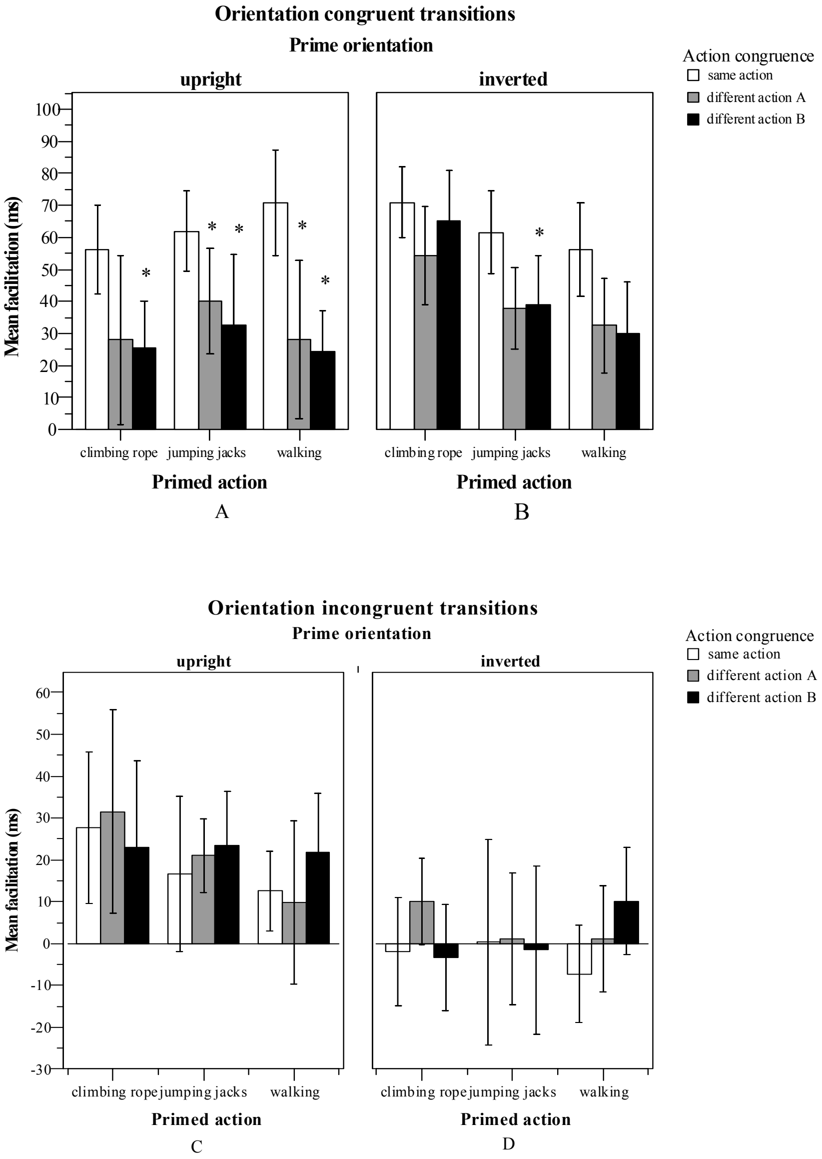


Figure 7.4. Mean priming effect for the orientation decision as a function of orientation congruence, prime orientation, primed action and action congruence. Error bars represent 95% confidence intervals. A star (*) represents a significant Bonferroni adjusted priming level difference between an action priming itself and when it was primed by a different action.

Another effect is the similar levels of priming for congruent upright transitions (Panel A) as well as for congruent inverted transitions (Panel B). Thirdly, there appears to be differences in priming levels for incongruent transitions, i.e., between Panels C and D. Upright primes seem to be able to prime inverted actions, whereas inverted primes have little influence on the upright oriented actions.

In order to statistically assess the main effects and interactions, a 2 (Orientation congruence: congruent vs. incongruent) x 2 (Prime orientation: upright vs. inverted) x 3 (primed action: climbing a rope vs. jumping jacks vs. walking) x 3 (Action congruence: same vs. different_A vs. different_B) repeated measures ANOVA was performed on the mean priming RTs for each of the 36 transitions. The designations different_A and different_B refer to the fact that incongruent action combinations consist of two categorically different actions. See Table 7.1 for the different combinations. Since the main emphasis is on the potential differences between priming effects for actions when they prime themselves and when they are primed by different actions, it is not necessary to specify the exact incongruent action combinations in every instance of the condition. The important fact is that the incongruent combinations consist of actions from different categories.

Main Effects

There was a main effect of orientation congruence, i.e., significantly greater overall priming for transitions of congruently oriented priming and primed actions than for incongruently presented actions (45 ms vs. 11 ms), $F(1,7) = 102.13$, $MSE = 832$, $\eta^2 = .94$, $p <.0001$. Priming levels for both conditions, however, were significantly greater than zero, $t(143) = 22.06$, $p < .0001$ for congruent transitions and $t(143) = 6.05$, $p < .0001$ for the incongruent transitions. The main effect of prime orientation was not significant, $F(1,7) = 3.73$, $MSE = 582$, $\eta^2 = .35$, $p = .095$. There was no priming difference between transitions beginning with upright primes and transitions beginning with inverted primes (31 ms upright vs. 25 ms inverted). Priming levels in both conditions, however, were significant greater than zero, $t(143) = 14.44$, $p < .0001$ for upright primes and $t(143) = 9.65$, $p < .0001$ for the inverted primes.

There were also significant differences between the levels of priming for the different primed actions, $F(2,14) = 4.75$, $MSE = 327$, $\eta^2 = .40$, $p = .027$. The mean overall levels of priming for the actions were as follows: climbing a rope 32 ms, jumping jacks 28 ms and walking 24 ms. None of the post-hoc Bonferroni adjusted multiple comparisons, however, reached significance. The main effect of action congruence was significant, $F(2,14) = 12.73$, $MSE = 309$, $\eta^2 = .65$, $p = .001$. This effect tests for the priming difference between transitions where the priming and primed action are the same action (e.g., walking) and where the priming and primed actions are different. Multiple post-hoc Bonferroni adjusted comparisons showed that there was significantly greater priming when the priming and primed actions were the same than when they were different, same 35 ms vs. different_A 25 ms ($t(7) = 4.04$, $p =$

.015) and same vs. different_B 24 ms ($t(7) = 3.994$, $p = .017$). There was no significant difference between different_A and different_B. Interpreting the meaning of the main effects is constrained by the possible interactions that the different variables enter into. The next section will present the analyses of the interaction effects.

Interaction Effects

In an experiment with 4 independent variables, there are 11 possible different interaction effects. In what follows, I will restrict my presentation of the interaction effects to the most theoretically relevant findings. Firstly, 6 of the 11 possible different interaction effects were statistically significant. Two-way interactions will be presented first, followed by 3-way interactions and then the 4-way interaction.

Despite the lack of a significant main effect of prime orientation, the 2-way interaction between *prime orientation* and *orientation congruence* was significant, $F(1,7) = 10.17$, $MSE = 1458$, $\eta^2 = .59$, $p = .015$. The effect of prime orientation differed reliably as a function of the levels of orientation congruence. The priming difference was greater between upright and inverted primes when transitions were *orientation incongruent*, i.e., for Panels C and D. Upright oriented primes led to an overall priming effect of 21 ms compared to a priming effect of 1 ms for incongruent transitions where the priming action was inverted. When an inverted action preceded an upright action, there was virtually no priming. This pattern was quite different when priming and primed actions had the same orientation. When transitions were *orientation congruent* and the priming actions were presented upright (Panel A), the priming effect was 41 ms compared to 50 ms for orientation congruent transitions and inverted priming actions (Panel B). There was considerable priming for *orientation congruent* actions for both upright and inverted primes. For orientation incongruent actions, it appears that upright actions can prime inverted actions but inverted actions could not prime upright actions. In contrast to results from previous research, the interaction here is quite different and shows both that inverted can prime inverted and upright can prime inverted *when* subjects are engaged in an orientation decision task.

The interaction between *orientation congruence* and *primed action* was not significant, $F<1$. Priming for the different primed actions, climbing rope, jumping jacks and walking, did not vary as a function of the levels of orientation congruence. This means that the *differences* in priming associated with the different actions mentioned above for the main effect of primed action were the same regardless of whether the transitions were orientation congruent or orientation incongruent. All of the three primed actions seem to be similarly affected by orientation congruence between priming and primed actions. There are no significant 2-way interactions between *primed action* and any other variable, and they will not be further discussed.

The interaction between *orientation congruence* and *action congruence* was significant, $F(2,14) = 27.82$, $MSE = 272$, $\eta^2 = .80$, $p = .00001$. The effect of action congruence varied as a function of the different levels of orientation congruence.

when priming and primed actions are orientation congruent, actions are better at priming themselves than when different actions are included in the priming-primed transitions (means: same action 63 ms vs. 37 ms and 36 ms for different$_A$ and different$_B$ respectively). The means show a quite different trend when the priming and primed orientations are different, same action 8 ms, different$_A$, 12 ms, different$_B$ 12 ms. This seems to show that subjects are able to make simple categorical discriminations between same and different actions when the priming and primed actions have the same orientation and that this ability disappears when priming and primed actions have a different orientation.

As indicated by the significant interaction between *prime orientation* and *action congruence*, $F(2,14) = 6.58$, $MSE = 86$, $\eta^2 = .48$, $p = .01$, the difference between the same and different actions in a priming transition varies as function of whether the primes were presented upright or inverted. When the primes were presented in an upright orientation, there was considerably more priming for same actions than for different actions, (means: same 41 ms, different$_A$ 26 ms and different$_B$ 25 ms). When, however, the primes were inverted, these differences were less (means same 30 ms, different$_A$ 23 ms and different$_B$ 23 ms). It appears that similar to the interaction between orientation and action congruence actions there is more priming for actions when they prime themselves than when they prime other actions, and this holds for upright presented primes but is diminished when primes are inverted.

The significant interaction between *prime orientation* and *action congruence* significantly interacts also with *orientation congruence*, $F(2,14) = 5.57$, $MSE = 77$, $\eta^2 = .44$, $p = .017$. This 3-way interaction is directly related to the hypotheses discussed at the beginning of this chapter. For orientation congruent transitions, there appears to be a difference between upright presented primes and inverted primes. When primes were presented upright, the difference between actions priming themselves (congruence) and when they prime different actions was greater than when the primes were inverted. This is evident by comparing the results in Panel A with the results in Panel B in Figure 7.4. Five out of 6 possible comparisons showed that there was significantly more priming for actions that primed themselves compared to when they were primed by other actions. Subjects seem to be able make categorical distinctions to a greater extent when the displays are orientation congruent and the primes are in an upright orientation. For orientation congruent inverted displays, only 1 out of 6 comparisons was significant.

Another aspect of the interaction between *prime orientation*, *action congruence* and *orientation congruence* concerns the differences between priming for “same” and highest “different” for congruent upright primes and congruent inverted primes, which are 31 ms vs. 18 ms respectively. The pattern of results for orientation incongruent transitions (Panels C and D) is quite different. It appears that subjects are unable to make simple category discriminations when the transitions consist of orientation incongruent displays, regardless of whether or not the prime orientations are upright or inverted. Differences between priming for the same and the highest “different” for upright oriented primes and inverted primes when the orientations are incongruent are -4 ms vs. -7 ms respectively. So, there appears to be an effect of prime orientation for congruent displays but not for incongruent displays. There is greater category discrimination in terms of priming differences when an upright display primes an upright display than when an inverted display primes an inverted display. This pattern holds for orientation congruent displays but not for orientation incongruent displays, as previously discussed within the context of the interaction between orientation congruence and action congruence.

The 3-way interaction between *orientation congruence*, *primed action* and *action congruence* was also significant, $F(4,28) = 6.28$, $MSE = 115$, $\eta^2 = .47$, $p = .001$. Recall that the *orientation congruence* x *action congruence* interaction was significant, which meant that subjects were able to make simple categorical discriminations only when the displays were orientation congruent. The addition of *primed action* as a significantly interacting variable indicates that the ability to make categorical discriminations when displays were orientation congruent depends on the primed action. For example, the difference in priming for walking when it primes itself and when the other actions prime it is 33 ms for the nearest different action. The comparable differences for jumping jacks and climbing rope are 23 ms and 22 ms respectively. This pattern of differences for the primed actions is quite different when the transitions consisted of orientation incongruent displays, where there are no positive values for priming effects.

Lastly, it should be mentioned that the 4-way interaction between orientation congruence, prime orientation, primed action and action congruence was not significant, $F<1$. The differences between priming as a function of orientation congruence, primed action and action congruence did not differ according the different levels of prime orientation. For orientation congruent displays, upright and inverted primes led to the same *patterns* of priming for the combinations of the different actions and levels of action congruence.

7.3 Discussion and Conclusions

Overall, the raw reaction times for both inverted and upright displays are somewhat faster than the reaction times that Verfaillie (1993) obtained in his experiments, which was around 500 ms. Verfaillie’s task was also different. If previous results from recognition and priming studies using static objects are compared with the results from the orientation decision about a dynamic display, we find that the times are roughly the same. For example, the results from Boucart and Humphreys (1992) show that subjects are performing a matching task using fragmented static object forms in about 500 ms. VanRullen and Thorpe (2001) investigated ultra-rapid visual categorization for animals and means-of-transportation and obtained reaction times of approximately 367 ms. If the time for the neural processing of the motor commands are subtracted from the reaction times, then correct responses take about 150 ms, which is also in line with the previously mention ERP-data (Jokish et al., 2005). Subjects performed the orientation decision task very quickly, which suggests that it requires relatively little cognitive effort, although subjects did report that the task was tiring after completing so many trials.

The results from this study demonstrate two different aspects of the inversion effect. Firstly, the raw reaction times to make an orientation decision (Table 7.2) showed that inverted displays took more time. This would be expected on the basis of results from previous studies mentioned in Chapter 6. The other aspect, which is unique to the findings for the repetition priming, showed that subjects were sensitive to differences between action categories when an upright display primed an upright display. When the prime was inverted, however, there was little processing that distinguished between the different actions. This is evident in the effect of action congruence in Panel A compared to Panel B. The pattern of priming effects in Panel B, however, is somewhat similar to the pattern in Panel A. Despite the lack of *significant* differences between an action when it primed itself and when it was primed by the other two actions, there is an evident trend in that direction, at least for jumping jacks and walking. One explanation for this finding is that subjects have learned to see inverted displays during the many trials in the experiment (Grossman, Blake & Kim, 2004; Jastorff, Koutzi & Geise, 2002). Palmeri and Gauthier (2004) documented significant learning of their Greeble stimuli and suggested that object (Greeble) identity may be automatically activated by expertise. As people become experienced at visually discriminating objects, access to knowledge mediating identification becomes more automatic. In Chapter 6 (6.4.1) I mentioned that Hiris et al. (2005) found that significant learning could occur for inverted biological motion stimuli when subjects were given the task of detecting inverted stimuli, and detection could be successfully performed by only focusing on the motion of a few dots. This ability was impaired when the task required a more global processing strategy.

There may be two different processes that occur when deciding whether or not the stimulus is inverted or upright. The first, as mentioned previously, is that it may only be necessary to visually process the patch-light figure as an *object*, i.e., a human being. This may be what is happening to a large extent in the *inverted* orientation congruent transitions (Panel B) and the orientation incongruent transitions in Panel C. For the upright orientation congruent transitions, there is clear evidence that there is additional processing that has access to the action performed by the patch-light figure. The evidence is in the significant 3-way interaction between *orientation congruence*, *prime orientation* and *action congruence*.

Given the many findings of orientation specificity of biological motion (e.g., Dittrich, 1993; Pavlova & Sokolov, 2000), the finding of similar levels of priming for Panel A and B of Figure 7.4 was unexpected. This is the first study to show that such levels of priming can be obtained with inverted displays of biological motion.

Another unexpected finding was the level of priming associated with the transitions in Panel C. Viewing an upright display led to significant levels of priming for all but 2 of the incongruent transitions. This finding contradicts, to some extent, previous findings of orientation specificity. What aspect of processing upright displays of biological motion facilitates the orientation decision for an inverted display? In beginning to answer this question, we need to look at the results in Panel D, where there is no priming effect. In this case, viewing an inverted prime was no better than viewing a neutral prime. If learning can account for the priming effect in Panel B, then learning clearly has not reached a level that leads to any processing advantage for upright displays when they are primed by an inverted display. The priming effect in Panel C could be due to an asymmetry in the activation of information associated with an action. If an upright action leads to activation of category information, which includes information about the human figure, then it may also activate information about different orientations of the human body. Information about possible different orientations of the human body is not a part of visually processing neutral stimuli. It does not appear that this information about the human body has any effect on being able to distinguish between different kinds of actions. There were no priming differences that occurred when actions primed themselves in comparison to when they were primed by other actions. The upshot is that the upright displays convey sufficiently enough information about the possible orientations of a human figure to create a priming effect.

Along similar lines, the lack of a priming effect in Panel D could be due to the relative differences for upright, inverted and neutral displays in activating information about the possible orientations of a human figure. Viewing an inverted display does not seem to *sufficiently* activate information about the other orientations of the human figure such that priming would occur for upright displays. The neutral stimuli contained no information about human bodies, but that does not mean that inverted displays do not contain any information about human bodies. The significant priming effects in Panel B suggest that inverted displays convey some information about the orientation of subsequent inverted displays, which indicates a processing advantage in relation to the neutral primes. This is admittedly speculation and needs to be systematically investigated. One way of testing for the effects of learning on incongruent priming would be to only present the priming stimuli one time during a study phase and then test for possible priming in a later test phase.

Previous findings suggest that high-level access to categorical information for inverted displays of biological motion is impaired relative to upright displays. The further claim here is that this access in turn is a result of limited configural or global level processing for inverted displays. Boucart and Humphreys (1992) showed this to be the case for static objects. Two major predictions were formulated to evaluate the role of categorical information in biological motion perception. First, if access to categorical information is automatic, then we should see categorical effects in priming for a task that does not require access to categorical information. Secondly, if this access is greater for upright than for inverted actions, we should see greater effects of action congruence for upright than for inverted actions.

Taken together, the results show a clear interaction between display orientation and access to categorical level information. The role of access to categorical level information is supported by Giese and Poggio's (2003) computational model of the recognition of biological movement. High-level areas in the form and motion pathways are selective for body shapes and specific human actions like walking and running. Feedforward processing from 'lower' visual areas along the different pathways activates motion pattern neurons that selectively encode motion patterns of human movement. Results from simulations of their model are consistent with the categorical processing of different actions based on psychophysical data.

In addition to the theoretical and modeling framework proposed by Giese and Poggio, the findings also suggest that the visual processing of upright displays is indicative of vision at a glance, whereas the processing implicated in viewing inverted displays indicates vision with scrutiny. Within the framework of Reverse Hierarchy Theory (RHT), access to categorical level information for upright displays is fast and automatic. This indicates that subjects had early access to high-level stored representations of human motion patterns that depicted specific actions.

In contrast to upright actions, orientation decisions for inverted actions took significantly more time and led to relatively less categorical level priming. In terms of RHT, this suggests that the feedforward mechanisms involved in the visual processing of inverted actions do not have the same level of access to stored representations of human motion patterns. The longer processing time for inverted displays could reflect the operation of feedback mechanisms that attempt to rebind local motion components (e.g., local rigidity) into a hierarchical whole for the purpose of identification. This is not to say that inverted displays cannot be reliably detected or recognized. It is rather a relative lack of access to categorical level information that distinguishes the processing of upright actions from inverted actions.

Chapter 8 - Conclusions and Future Work

The purpose of this chapter is to put the results of the previous chapters into a larger theoretical perspective and to discuss the implications of the results for future research. I will also discuss methodological limitations of the research presented here. Before moving on to these important aspects, I will briefly discuss the main findings.

8.1 Main Findings

The general results from the empirical studies further support the idea that action concepts contain information about the spatiotemporal dynamic form of actions, and that this high level configural information is used by people to group action exemplars and structure action categories. The results from the experiments in Chapters 3 and 5 represent converging evidence for prototype effects for categories of natural actions. When presented with the task of listing verbs that name actions that could easily be recognized when seen and could be visualized as a mental image, subjects produced lists which, when analyzed, revealed a structure that supported the notion of basic level and subordinate level actions. This was the case for both American English and Swedish speaking samples. The relationship between basic level and subordinate level actions indicates that action categories have graded structure around an action prototype and that this cognitive organization is cross-linguistically similar.

The experiments in Chapter 5 were conducted independently of one another and directly investigated the extent to which subjects produced different typicality ratings and verification times for different action exemplars in relation to a category label. Action exemplars were presented as patch-light displays of biological motion. The results showed that typicality ratings can be used as reliable predictors of verification reaction times. Even though the verification task did not explicitly require subjects to determine the typicality of an action exemplar, verification reaction times reflected the prototype structure of action categories. The spatiotemporal relatedness of action exemplars led to a context effect. The greater the spatiotemporal relatedness between an action exemplar and a contrast category label, the longer it took subjects to respond in the verification task. Subjects also made more errors when action exemplars and contrast categories were perceptually related. It is important to note, however, that subjects were still able to make clear categorical distinctions between most action exemplars and contrast category labels. This point will be discussed in greater detail shortly.

The results showing the categorical distinctions between actions appear to be in conflict with the following claim of Vinson and Vigliocco (2002):

> We discuss impairments and organization in terms of semantic fields rather than categories, because category-level distinctions are far less clear for actions than for objects; superordinate category labels for actions are unclear, category boundaries among actions are vague or nonexistent, and 'category-specific' impairments have not generally been observed within the general class of actions. 'Semantic field' is thus used as a general term to refer to groups of words that are organized according to meaning. (p.318, footnote 1)

The results from the typicality and verification studies in Chapter 5 indicate that subjects can and do indeed make category level distinctions. Recall that 55 out of 60 conditions of typicality judgments for non-matching category labels and actions resulted in values of less than 1. This meant that most subjects were judging the actions as "not at all typical." Furthermore, this was the case for actions that were to some extent perceptually similar, for example, when determining how typical a running action was of the category of kicking. The exception to this trend was when running or waving instances were judged as being at least "somewhat typical" of RUNNING or WAVING contrast categories. There may be a general difference between the overall conceptual structure for nouns and verbs, but the class of verbs relating to natural actions might exhibit closer ties to object categories, and in this sense exhibit relatively clear category boundaries.

Results from the repetition priming experiment in Chapter 7 indicated that high level configural information can be implicitly activated, i.e., primed. This information appears to be used to make category distinctions between different actions presented as patch-light displays of biological motion. Access to the configural information, however, is limited to upright oriented displays or requires learning in the case of inverted displays. Previous findings of the orientation specificity of biological motion perception need to be somewhat revised given the results showing that upright displays can indeed prime inverted displays when subjects are instructed to make a decision about the display orientation. On the other hand, inverted displays do not facilitate the orientation decision for upright displays. The priming effects are asymmetrical.

Recall that previous findings have shown that although inversion significantly disrupts visual processing of biological motion displays, it does not always lead to a complete lack of identification or recognition (Grossman & Blake, 2001; Pinto & Shiffrar, 1999). The results of Grossman and Blake (2001) showed that viewing inverted displays led to a significant reduction of brain activity compared to upright displays, but the activity levels associated with the viewing of inverted displays remained significantly above levels associated with viewing scrambled displays. How does this result fit in with the obtained priming results from Chapter 7? No significant levels of priming were obtained for the inverted-upright (orientation incongruent) transitions (Panel D, Figure 7.4). This does not mean that subjects were completely unable to visually process the inverted displays. It simply means that inverted displays had no more of a facilitating effect than neutral displays on the orientation decision for the immediately following upright displays.

8.1.1 Disclaimers

The work here concerning the categorization of human actions has not focused on finding *the* basic level for action categories. The emphasis has rather been on investigating how we might organize our knowledge about the actions of others. The idea that there might be a basic level is just one way of viewing how knowledge of human actions might be categorically organized. I doubt that a hierarchical view of the organization of knowledge about human actions in terms of superordinate, basic and subordinate level categories will provide us with a complete framework for understanding how we organize our knowledge about actions. But it is a valuable starting point with an established research record of value for starting to look at issues regarding the organization of knowledge about human actions.

Furthermore, the experiments here have not investigated the memory systems involved in action perception. It remains a possibility that categorical effects are due to the demands of current task and formed within working memory, rather than being structures in long-term memory. To my knowledge, however, no research within biological motion processing has specifically addressed this issue. I think it is rather the case that most researchers would assert that biological motion perception relies on access to information stored in long-term memory. For example, the highest level in the Giese and Poggio (2003) model includes complete motion sequences such as walking, throwing, running, etc. The motion sequences are encoded by motion pattern neurons. These motion pattern neurons seem to be good candidates for long-term memories of natural actions.

8.2 Methodological Limitations

One methodological limitation has to do with the number of subjects, or number of measurements that figured into the statistical analyses in the experiments. For example, the cross-linguistic analyses in Chapter 3 were based on 39 subjects in each language group. The MDS analyses were based on all possible verb pairs for the most frequent verbs. However, just because a verb occurs frequently, does not mean that it will frequently occur *together* with *all other* frequent verbs. Some verb *pairs*, only occurred infrequently, and some did not occur at all on the lists. This could lead to a problem of interpretation for the Euclidean distances associated with those verb pairs.

One way of dealing with this limitation is to focus on the verb pairs that occurred more frequently. The best method of dealing with this limitation in future studies is to test about twice as many subjects in the two language groups. The effect of this will be to decrease the error variance and possible effects due to outliers. Why did I not include more subjects in the studies? One reason was that I was able to obtain results that were similar to the American English sample by only using 39 subjects in the Swedish speaking group. A second reason had to do with the time that was required to extract the total frequency data and obtain the data for the calculation of the mean ordinal positions as well as check each list for all possible verb combinations. This procedure was very time consuming. In order to make the data extraction more efficient, the procedures should be done on a computer. A computer program could easily register and sort data by frequency, mean ordinal position and verb pair ordinal distance. I will develop this idea below.

As mentioned in section 5.2.2.2, increasing the number of subjects in the typicality and verification experiments would have led to more stable means for the different conditions. The standard deviations for the verification reaction times were quite large. This made it difficult to obtain statistically significant effects for small differences between condition means. The results did in fact show a strong relationship between typicality ratings and verification times in the typicality-RT effect, but there was only a partial statistical confirmation of the prototype structure that was obtained with the typicality ratings. In terms of null hypothesis testing, increasing the number of subjects would likely lead to an increase in power and therefore increase the probability of rejecting a false null hypothesis. Despite the need for greater power, there is reason to assert that a prototype structure could be obtained for all four categories, i.e., *running*, *kicking*, *throwing* and *waving*. Recall that the typicality ratings led to a three-tiered graded structure for *running* and *kicking* and a two-tiered graded structure for *throwing* and *waving*. However, in the verification study, no significant differences were found between the different kicking exemplars, and only a two-tiered graded structure was found for *running*. The correlation showing the typicality-RT effect shows however that typicality ratings and verification RT are strongly related and thereby indicate a clear prototype structure for all of the action categories. The conclusion that I reach here is that there is sufficient data to support the hypothesis of a prototype structure for, at least, a limited domain of action categories.

A further methodological limitation is the limited number of action categories used in the studies. The actions used here constituted basic kinds of natural actions that are likely found in most cultures and language groups. It might be the case that while prototype structures exist for these kinds of actions, other, more cultural and context determined, actions may not exhibit prototype effects. A question arises as to the extent to which the current results apply to other kinds of action or action categories. The purpose of the experiments in this book has been to determine whether or not prototype effects can be obtained for action categories, which I think has been shown. Indeed, an issue for further research might be to determine possible limitations of prototype effects for action categories. The findings of Giese and Lappe (2002) and the more recent findings of Giese, Thornton and Edelman (2008) are certainly consistent with the results from my experiments.

8.3 Proposals for Future Research

When analyzing and discussing the experimental results from the previous chapters, a number of follow-up questions and ideas for further experiments have arisen. In this section, I will present some of those ideas and discuss their relevance for advancing our knowledge of the perception of biological motion. It is important to note that the proposals vary in the extent to which they are developed. Instead of providing experimental details, the purpose is to point to directions in which further research can address some of the remaining issues. The proposals will also be discussed in relation to language, categorization and the perception of actions.

8.3.1 Action Naming, the Basic level and Orientation Specificity

As mentioned in Chapter 2, objects tend to be identified at the basic level. When people are presented with pictures of objects, they tend to use the same labels for the objects. Further converging evidence for a basic level for action categories could be obtained by letting subjects identify action exemplars presented as point-light displays. For example, the same actions from the typicality and verification experiments could be used in an identification experiment. The spoken name of the actions would be recorded as well as the time taken to identify them. One prediction of the identification phase would be the occurrence of verbal descriptions of the actions as they become more “distant” from the prototype and perhaps more similar to exemplars from contrast categories. The more similar an action exemplar is to exemplars from contrast categories, the more important it may be to identify the action by lexically marking the closeness of the exemplar to contrast categories. I suspect, however, that there may be a greater tendency for this to occur with action categories that have similar spatiotemporal patterns (Giese & Lappe, 2002).

A further manipulation could include the orientation of the actions. The purpose of the orientation manipulation would be to investigate the extent to which action naming might change due to the orientation of the action. Recall that Sumi (1984) originally reported that some subjects were able to see a human body but failed to see it as being upside down. Subjects described the movements of arms and legs but also failed to see any meaningful coherent human action. By including the orientation manipulation together with the different action exemplars, it is possible to systematically assess the effect of stimulus orientation on naming. Inverted displays can serve as kind of benchmark against which naming performance can be compared.

8.3.1.1 Long-term Priming

Results from the experiment in Chapter 7 showed that implicit activation of categorical information about the actions led to a significant facilitation of the orientation decision time. A further issue concerns the extent to which this implicit activation would also facilitate an explicit categorization task. The orientation decision did not require an explicit response about the category membership of the action exemplars, and it is therefore difficult to draw any verifiable conclusions about the effects of the activation of category knowledge on performance in an *explicit* task. If, however, subjects are given an explicit task such as action identification, and performance is facilitated by previous exposure to actions, this would indicate that the priming action activates categorical information. A proposal for studying the effect of implicit processing on explicit responses would be to expose subjects to a number of actions, also presented as patch-light displays, and then in a later test phase measure the effect of previous exposure to the actions by measuring explicit identification performance. This could be done by using an old-new priming paradigm. The basic gist here would be to construct the identification phase such that subjects would be asked to identify actions that were previously presented in the study phase, i.e., the old actions, and identify actions that were not previously presented in the study phase, i.e., new actions. If previous exposure to actions in the study phase facilitates identification, then significantly more “old” actions should be identified than “new” actions. Of course, counterbalancing would have to be used to ensure that all actions would appear an equal number of times in the study and test phases.

In order to test the effects of orientation on action identification and as a partial replication of the results presented in Chapter 7, orientation congruence could be manipulated between the study and test phase of the experiment. If categorical information is activated during the study phase and it facilitates later performance on the identification task, then we should see significantly better identification for upright congruently presented actions in the study and test phases compared with action identification for “new” actions in the test phase. This would replicate the pattern of results shown in Panel A in Figure 7.4. For the condition where inverted actions are presented in both the study and test phases, there is reason to believe that no priming effect would occur. If the levels of priming for the inverted-inverted transitions in Panel B in Figure 7.4 are the result of learning over many trials, then just one exposure to inverted displays in a study phase will not likely be sufficient to facilitate the identification of actions. In the case of upright displays being presented in the study phase and inverted displays being presented in the test phase, it may be the case that access to categorical information also spreads to inverted actions which will then facilitate action identification for the inverted displays. If, however, the obtained priming effects in Panel C in Figure 7.4 are due to learning over many trials, there should be no priming effect when subjects are only exposed to one presentation of the actions in a study phase. Consistent with the effects in Panel D in Figure 7.4, I would not expect any priming effects for inverted actions presented in the study phase and then presented upright in the identification phase. With this proposed experiment, it is possible to follow up the results from the repetition priming experiment and to test further hypotheses about the effect of implicit access to categorical information on explicit responses.

8.3.1.2 Attention, Automatic Processing and Pop-Out

None of the experiments presented in this book have specifically investigated the role of attention in the processing of biological motion. As mentioned in Chapter 6 (section 6.5), attention can play different roles depending on the level of processing needed to complete a given task when viewing point-light displays. The proposals in this section present different ways of building on previous findings about attention and biological motion processing.

The phenomenal visual experience of viewing point-light display seems to have a pop-out-like quality. This means that the coherent configural arrangement of the motion of the dots in a point-light display leads to a rather immediate impression of the depicted action. Strictly speaking, however, as a phenomenon of attention, pop-out is something that occurs during visual search, and is relatively unaffected by the number of co-occurring distratctors (e.g., Ahissar & Hochstein, 2004; Hochstein & Ahissar, 2002; Treisman & Gelade, 1980; Wolf & Horowitz, 2004). It should be noted, howver, that different kinds of co-occurring distractors can place different attentional demand on the visual processing of biological motion (Hunt & Halper, 2008). The "automatic" processing of displays of biological motion is also closely tied to the notion of attention and means that relatively little attentional resources are required to perform a given task. In this sense then, object pop-out can be obtained with minimal demands placed on attention. In contrast, visual search tasks that require focused attention, i.e., conjunctive feature searches, do not lead to a pop-out-like visual experience. Focused attention usually involves a top-down driven control process and is therefore a controlled process rather than an automatic process.

Results from Thornton et al. (2002) provided an indication of automatic processing of biological motion. When point-light displays were masked by randomly moving points,[45] performance on a secondary task was relatively unaffected in relation to a baseline task. However, when a scrambled mask was used, performance on the secondary task dropped to chance levels. This indicates that the visual processing of biological motion in the context of a random mask is largely automatic. For the scrambled masks, focused attention was apparently required to correctly

---

[45] In a random mask, the positions of masking points are randomly determined. In addition, the trajectories of the randomly positioned points are random. This is in contrast to a scrambled mask in which the trajectories of the randomly positioned points are copied from the target point-light figure.

complete the task, which was why attentional resources could not be spared for the secondary task.

In order to further test for the pop-out-like quality of the visual processing of biological motion, I suggest a further manipulation of the kinds of distractor elements that are used to construct the scrambled masks (cf. Cutting et al., 1988; Hiris, 2007). The scrambled masks used in previous detection experiments have been created by duplicating the individual trajectories of the points from the *target* point-light figure. For example, if the target is a point-light walker, then the scrambled mask contains moving points from the point-light walker. If the scrambled mask is composed of masking elements that come from *different* (basic level) actions, detection may be easier since there is even less motion overlap between the motion of the masking elements and the target. Since the motions of the masking elements still come from a human actor, detection performance may not be as good as when the masking elements are purely random. In addition to mixing targets and masking elements from different actions, the trajectories from the individual points of inverted displays could be used to create scrambled masks. If inversion of whole point-light figures disrupts visual processing, then scrambling the trajectories of the individual inverted points would likely lead to very little, if any, reduction in detection of an upright target compared to detection of an upright target in a completely random mask. Detection of a target in a completely random mask would function as a baseline control condition.

Another experiment using the visual search paradigm could be to include actions from different categories as targets among distractors from contrast categories. Instead of embedding the target in a dynamic mask, the distractors would consist of whole point-light figures (Thornton & Vuong, 2004). The idea here is that visual search performance for a target action among different basic level actions should be achieved with a higher degree of automaticity (quickness, accuracy and relatively low levels of attention) than a visual search that requires the subject to detect a target action among subordinate level actions. Perhaps neutral stimuli of the kind used in Chapter 7 could be used as a baseline condition. Another manipulation would be to include the factor of orientation for the target as well as the distractor actions which could be done with the flanker-interference paradigm used in Thornton and Vuong (2004) where they showed that upright displays are processed incidentally in a flanking task. If, as I claim, inverted displays impair access to semantic level information, then it should be the case that inference effects should be less with inverted displays than with upright displays. Predictions about the effects of orientation and action congruence depend crucially on the task. Within this experimental setup, it is possible to use the orientation decision task, i.e., to determine whether the target is upright or inverted. Given this task, any interference due to the flankers would likely be the result of incidental processing.

8.3.1.3 Cross-linguistic Studies of Lexicalized Action Categories

The cross-linguistic studies in Chapter 3 provide a good empirical basis from which to further investigate the potential similarities of semantic spaces for lexicalized action verbs in different languages. In addition to looking at the relationships between different frequencies for basic level and subordinate level verbs, I intend to focus more on the extent to which similarly constrained human natural actions are reflected in the naming patterns of different languages. The recent findings of Malt et al. (2008) support the view that despite differences in linguistic typology for expressing motion events (Slobin, 2004) subjects tend to use the same categorical naming patterns for instances of human locomotion, namely running and walking. In this case, action naming appears to reflect the structural discontinuity between walking and running. This is consistent with an embodied language perspective. It is also a line of research that I and my colleagues have already begun to explore by recruiting subjects from different language communities to create lists of verbs according to the instructions in Chapter 3.[46]

As mentioned in Chapter 1, there is accumulating evidence showing that our understanding for verbs that name the actions of various body parts like arms, legs, hands, etc. is highly correlated with neural activity in premotor areas that correspond to watching actions that involve those body parts (Aziz-Zadeh, Wilson, Rizzolatti & Iacoboni, 2006; Tettamanti, Buccino, Saccuman et al., 2005). In this sense, the semantic organization of verbs for concrete bodily motions may be determined to some extent by the somatotopic activation of neurons for the different body parts. The important notion here is that multidimensional scaling (MDS) could perhaps be used as a tool to (re)construct the relative distances in the somatotopic organization of the body parts along the premotor cortex. This should not be taken to mean that the linguistic data will tightly match the organization of neurons in the premotor cortex. One reason this is likely not the case is the involvement of other differently organized cortical areas in action identification, e.g., the STS, the premotor cortex (Saygin et al., 2004; Tai, Scherfler, Books, Sawamoto & Castiello, 2004) and parietal areas (Battelli et al., 2003). Support for an embodied language perspective can be obtained by finding a similar organization across different languages for the bodily motion verbs and verbs for vocal and mouth movements mentioned in Chapter 3. If a similar organization exists across very different languages, it would further suggest a common embodied basis for at least some of the verbs that name natural actions.

As mentioned in the methodological limitations above, future work within this area should include the development of a computerized data gathering system. This would allow for a more efficient collection of data and subsequent analyses. With

---

[46] Barbara Gawronska and I have extended this research to include Polish speaking subjects, and, together with Sofia Kasviki, we have gather verb data from Greek speaking subjects.

more subjects, more fine-grained analyses should be possible. These fined grained analyses would include being able to specifically determine the semantic space for all different kinds of kicking actions. Given the limitations of the number of subjects that we have tested thus far, there is not enough statistical material to do appropriate analyses on this level.

In addition to the findings in Chapter 3, it is also likely the case that there are other significant patterns of semantic associations between the verbs. Preliminary results (Hemeren & Gawronska, 2007; Hemeren, Kasviki & Gawronska, 2008) show that there is also a tendency among subjects to list actions that belong to a theme or script rather than according to a strict concept hierarchy. Previous findings for this kind of organization were discussed in Chapter 2, section 2.4.2. With experiments of this kind, we should be able to capture cross-linguistic regularities in the semantic organization of verbs for natural actions as well as capture some language and culture specific aspects.

8.3.1.4 The Role of Force Patterns

Given the role played by the *forces* involved in actions, more experiments need to be performed to investigate the effect that different force patterns have on the prototype structures of action categories (Gärdenfors, personal communication). If actions can be represented in terms of dimensions in a conceptual space, then more research needs to be done to find out what qualities are associated with the dimensions that structure the space. One shortcoming of the MDS-studies mentioned above is the fact that we do not know what the dimensions of the MDS-solutions stand for. Another way of investigating the conceptual spaces for actions is to systematically vary *known* dimensions and let subjects make similarity judgments between pairs of actions based on those dimensions. In this way, we should be able to examine clear category breaks between continuously varying dimensions, which would also provide us with evidence regarding the existence of psychological boundaries between action categories based on, for example, force patterns.

8.4 Concluding Comments

Human self-generated motion is inherently dynamic. The implications of this for theories of action recognition, concept formation and categorization is that action identification necessarily involves understanding the ways in which configurations of the human body can change over time. A crucial aspect of this understanding has to do with possessing knowledge about the physical and cultural/social constraints that limit the ways in which the body can move. Whereas the physical constraints have to do with the physical characteristics of the external environment, e.g., gravity and surface textures, and the biomechanics of our human bodies, the cultural/social constraints concern the kinds of movements that are accepted and encouraged in a community.

During a lifetime, we develop an expertise at identifying the actions of others. Our visual experience is rich with examples of people acting in different ways. We become experts at action perception. In contrast to the visual experience of observing the actions of others, the perception of our own actions is based on a completely different point of view (Jacobs, Pinto & Shiffrar, 2004; Jacobs & Shiffrar, 2005). We have a lot of visual experience at perceiving the actions others, and we have a lot of experience at perceiving our own actions, but the basis for the perceiving our own actions is not so much visual as it is proprioceptive or motor based. It also appears that our perception of the actions of others is usually guided by having motoric schemas or representations for the motion of our own body parts (de Vignemont, Tsakiris & Haggard, 2006). Remarkably, however, having a mental motoric schema *does not* depend on actually having the appropriate body part. For example, it is possible to develop appropriate motoric representations for a simple hand motion *without* access to an actual hand (Funk, Shiffrar & Brugger, 2005). This appears to be possible due to the presence of phantom sensations of congenitally missing limbs. Despite the lack of hands since birth, it is possible to experience vivid phantom postural and movement sensations in them (Brugger, 2006). This striking example is additional evidence of the connection between the visual identification of observed actions and the involvement of motor resonance with our own bodies (e.g., Calvo-Merino et al., 2006; De Maeght & Prinz, 2004; Hari et al., 1998; Jackson & Decety, 2004; Lozano et al., 2008; Viviani, 2002) and the body's role in language (e.g., Gibbs, 2003; Glenberg & Kaschak, 2002; Yeh & Barsalou, 2006; Zwan & Taylor, 2006).

Thinking (including perception) relies on our ability to organize information, and the organization of information is achieved through the use of concepts to categorize "objects" in our surroundings. Action concepts allow us to think about and understand the actions of others in terms of intentions and goals and in terms of our own abilities to purposefully move about. Through the activation of motor imagery, action concepts allow us to simulate dynamic situations and the potential consequences in them. To the extent that language and communication also use concepts, then concepts appear to be an important link between language and thought. My goal has been to suggest and empirically demonstrate a prototype structure for action categories and the effect this structure has on language via the semantic organization of verbs for natural actions. It has also been to show that implicit access to categorical information can affect the perceptual processing involved in making a simple decision about the orientation of an action. Therefore, given the growing emphasis on, and development of, interdisciplinary efforts in cognitive science, the themes of language, categorization and perception in this book represent a contribution to those efforts.

References

Aggarwal, J. K. & Cai, Q. (1999). Human motion analysis: A review. *Computer Vision and Image Understanding*, 73(3), 428-440.

Ahissar, M. & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8(10), 457-464.

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

Aktinson, A. P., Dittrich, W. H., Gemmell, A. J. & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, 33, 717-746.

Arbib, M. A. (2008). From grasp to language: Embodied concepts and the challenge of abstraction. *Journal of Physiology – Paris*, doi:10.1016/j.jphysparis.2008.01.001.

Aziz-Zadeh, L. & Damasio, A. (2008). Embodied semantics for actions: Findings from functional brain imaging. *Journal of Physiology*, doi:10.1016/j.jphysparis.2008.03.12.

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G. & Iaconboni, M. (2006). Congruent embodied representations of visually presented actions and linguistic phrases describing actions. *Current Biology*, 16, 1818-1823.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, 10, 82-93.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, 11, 211-227.

Barsalou, L. W. (1985). Ideals, Central Tendency, and Frequency of Instantiation as Determinants of Graded Structure in Categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(4), 629-654.

Barsalou, L. W. (1987). The instability of graded structure: implications for the nature of concepts. In U. Neisser (Ed.) *Concepts and Conceptual Development* (pp.101-140). Cambridge University Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-660.

Barsalou, L. W., Simmons, W. K., Barbey, A. K. & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 7(2), 84-91.

Battelli, L., Cavanagh, P. & Thornton, I. M. (2003). Perception of biological motion in parietal patients. *Neuropsychologia*, 41, 1808-1816.

Battig, W. F., & Montague, W. E. (1969). Category names for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 80(3), Part 2(June 1969), 1-46.

Beauchamp, M. S., Lee, K. E., Haxby, J. V. & Martin, A. (2003). fMRI responses to video and point-light displays of moving humans and manipulable objects. *Journal of Cognitive Neuroscience*, 15(7), 991-1001.

Beintema, J.A. & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Science of the United States of America*, 99, 5661-5663.

Berry, D. S., Kean, K. J., Misovich, S. J. & Baron, R. M. (1991). Quantized displays of human movement: A methodological alternative to the point-light display. *Journal of Nonverbal Behavior*, 15(2), 81-97.

Bertenthal, B. I. & Pinto, J. (1994). Global processing of biological motions. *Psychological Science*, 5, 221-225.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2), 115-147.

Blake, R. (1993). Cats perceive biological motion. *Psychological Science*, 4(1), 54-57.

Blake, R. & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, 58, 47-73.

Blakemore, S.-J. & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, 2, 561-567.

Blakemore, S.-J. & Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia*, 43, 260-267.

Bonin, P., Boyer, B., Méot, A., Fayol, M. & Droit, S. (2004). Psycholinguistic norms for action photographs in French and their relationships with spoken and written latencies. *Behavior Research Methods, Instruments, & Computers*, 36(1), 127-139.

Bosbach, S., Knoblich, G., Reed, C. L., Cole, J. & Prinz, W. (2006). Body inversion effect without body sense: Insights from deafferentation. *Neuropsychologia*, 44, 2950-2958.

Boucart, M. & Humphreys, G. W. (1992). Global shape cannot be attended without object identification. *Journal of Experimental Psychology: Human Perception and Performance*, 18(3), 785-806.

Boucart, M. et al. (2000). Automatic object identification: An fMRI study. *NeuroReport*, 11(11), 2379-2383.

Boutsen, L. & Humphreys, G. W. (2003). The effect of inversion on the encoding of normal and “thacherized” faces. *The Quarterly Journal of Experimental Psychology*, 56A(6), 955-975.

Brown, R. (1958). How Shall A Thing Be Called? *Psychological Review*, 65(1), 14-21.

Brugger, P. (2006). From phantom limb to phantom body. In G. Knoblich, I. M. Thornton, M. Grosjean & M. Shiffrar (Eds.) *Human body perception from the inside out* (pp. 171-209). Oxford: Oxford Univ. Press.

Buccino, G., Binkofski, F. & Riggio, L. (2004). The mirror neuron system and action recognition. *Brain and Language*, 89, 370-376.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G. & Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, 13, 400-404.

Calder, A.J. & Jansen, J. (2005). Configural coding of facial expressions: The impact of inversion and photographic negative. *Visual Cognition*, 12(3), 495-518.

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E. & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity an action observation. *Current Biology*, 16, 1905-1910.

Carey, S. & Diamond, R. (1994). Are faces conceived as configurations more by adults than by children? *Visual Cognition*, 1, 253-275.

Casey, P. J. & Heath, R. A. (1989). A semantic memory sentence verification model based on relative judgment theory. *Memory & Cognition*, 17(4), 463-473.

Casey, P. J. & Heath, R. A. (1990). Semantic memory retrieval: Deadlining the typicality effect. *The Quarterly Journal of Experimental Psychology*, 42A(4), 649-673.

Casile, A. & Giese, M. A. (2005). Critical features for the recognition of biological motion. *Journal of Vision*, 5, 348-360.

Casile, A. & Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Current Biology*, 16, 69-74.

Cavanagh, P., Labianca, A. T. & Thornton, I. M. (2001). Attention-based visual routines: sprites. *Cognition*, 80, 47-60.

Chang, D. H. F. & Troje, N. F. (2008). Perception of animacy and direction from local biological motion signals. *Journal of Vision*, 8(5):3, 1-10.

Chrisley, R. & Ziemke, T. (2002). Embodiment. In *Encyclopedia of Cognitive Science* (pp. 1102-1108). London: Macmillan Publishers.

Clark, A. (1999). An embodied cognitive science. *Trends in Cognitive Sciences*, 3(9), 345-351.

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences*, 10(8), 370-374.

Cliff, N. (1993). What is and isn't measurement. In G. Keren & C. Lewis (Eds.), *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues* (pp. 59-93). Hillsdale, NJ; Lawrence Erlbaum.

Collins, A. M. & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 9, 432-438.

Corter, J. E. & Gluck, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin*, 111(2), 291-303.

Cutting, J. E. (1981). Coding theory adapted to gait perception. *Journal of Experimental Psychology: Human Perception and Performance,* 7(1), 71-87.

Cutting, J. E., Moore, C. & Morrison, R. (1988). Masking the motions of human gait. *Perception & Psychophysics*, 44(4), 339-347.

Daems, A. & Verfaillie, K. (1999). Viewpoint-dependent priming effects in the perception of human actions and body postures. *Visual Cognition*, 6(6), 665-693.

Daprati, E., Wriessnegger, S. & Lacquaniti, F. (2007). Knowledge of one's kinematics improves perceptual discrimination. *Consciousness and Cognition*, 16, 178-188.

Dekeyser, M., Verfaillie, K. & Vanrie, J. (2002). Creating stimuli for the study of biological-motion perception. *Behavior Research Methods, Instruments, & Computers*, 34(3), 375-382.

De Maeght, S. & Prinz, W. (2004). Action induction through action observation. *Psychological Research*, 68, 97-114.

de Vignemont, F., Tsakiris, M. & Haggard, P. (2006). Body mereology. In G. Knoblich, I. M. Thornton, M. Grosjean & M. Shiffrar (Eds.) *Human body perception from the inside out* (pp. 147-170). Oxford: Oxford Univ. Press.

di Pellegrino, G., Fadiga, L. Fogassi, L, Gallese, V. & Rizzolatti, G. (1992). Understanding motor events. *Experimental Brain Research*, 91, 176-180.

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

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 (Eds.) *Gesture-Based Communication in Human – Computer Interaction* (pp. 3-22). Berlin: Springer-Verlag.

Dittrich, W. H. & Lea, S. E. G. (1994). Visual perception of intentional motion. *Perception*, 23, 253-268.

Dittrich, W. H., Lea, S. E. G., Barrett, J & Gurr, P. R. (1998). Categorisation of natural movements by pidgeons: Visual concept discrimination and biological motion. *Journal of the Experimental Analysis of Behavior*, 70, 281-299.

Dittrich, W. H., Troscianko,T., Lea, S. E. G. & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance, *Perception*, 25, 727-738.

Farah, M. J., Wilson, K. D., Drain, M. & Tanaka, J. N. (1998). What is ”special” about face perception? *Psychological Review*, 105(3), 482-498.

Fisher, C., Gleitman, H. & Gleitman, L. R. (1991). On the Semantic Content of Subcategorization Frames. *Cognitive Psychology*, 23, 321-392.

Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Clarendon Press.

Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics*, 33, 575-581.

Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, 94(4), 427-438.

Funk, M., Shiffrar, M. & Brugger, P. (2005). Hand movement observation by individuals born without hands: phantom limb experience constrains visual limb perception. *Experimental Brain Research*, 164, 341-346.

Gallese, V. & Lakoff, G. (2005). The brain’s concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22, 455-479.

Gärdenfors, P. (2000). *Conceptual Spaces*. Bradford Books, MIT Press.

Gärdenfors, P (2007). Representing actions and functional properties in conceptual spaces. In T. Ziemke, J. Zlatev & R. Frank (Eds.) *Body, language, and mind: Embodiment* (vol. 1, pp. 167-195). Berlin: Mounton de Gruter.

Garcia, J. O. & Grossman, E. D. (2008). Necessary but not sufficient: Motion perception is required for perceiving biological motion. *Vision Research*, doi:10.1016/ j.visres. 2008.01.027.

Gauthier, I. & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2), 431-446.

Gibbs, R. (2003). Embodied experience and linguistic meaning. *Brain and Language*, 84, 1-15.

Giese, M. A. (2002). Prototypes of biological movements in brains and machines. In H. H. Bülthoff, S. W. Lee, T. Poggio & C. Wallraven (Eds.) *Biologically motivated Computer Vision* (pp. 157-170). Berlin: Springer-Verlag.

Giese, M. A. (2006). Computational principles for the recognition of biological movements. In G. Knoblich, I. M. Thornton, M. Grosjean & M. Shiffrar (Eds.) *Human body perception from the inside out* (pp. 323-359). Oxford: Oxford Univ. Press.

Giese, M. A. & Lappe, M. (2002). Measurement of generalization fields for the recognition of biological motion. *Vision Research*, 38, 1847-1858.

Giese, M. A. and Poggio, T. (2002). *Biologically plausible neural model for the recognition of biological motion and actions*. AI Memo 2002-012, MIT – AI-lab.

Giese, M. A. & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, 4(3), 179-192.

Giese, M. A., Thornton, I. & Edelman, S. (2008). Metrics of the perception of body movement. *Journal of Vision*, 8(9):13, 1-18, http://journalofvision.org/8/9/13/, doi:10.1167/8.9.13.

Glenberg, A. M. & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558-565.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123(2), 178-200.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, 24, 608-628.

Goldstone, R. L. & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, 65, 231-262.

Graf, M., Reitzner, B., Corves, C., Casile, A., Giese, M. & Prinz, W. (2007). Predicting point-light actions in real-time. *NeuroImage*, 36, T22-T32.

Grill-Spector, K. & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science*, 16(2), 152-160.

Grossman, E. D. (2006). Evidence for a network of brain areas involved in perception of biological motion. In: G. Knoblich, I. M. Thornton, M. Grosjean & M. Shiffrar (Eds.) *Human body perception from the inside out* (pp. 361-384). Oxford: Oxford Univ. Press.

Grossman, E. D., Battelli, L. & Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Research*. 45, 2847-2853.

Grossman, E. D. & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research*, 41, 1475-1482.

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

Grossman, E. D., Blake, R. & Kim, C-Y. (2004). Learning to see biological motion: Brain activity parallels behavior. *Journal of Cognitive Neuroscience*, 16:9, 1669-1679.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27, 377-442.

Haggard, P. (2005). Conscious intention and motor cognition. *Trends in Cognitive Sciences*, 9(6), 290-295.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 18, 441-461.

Hampton J. A. (1995). Testing prototype theory of concepts. *Journal of Memory and Language*, 34, 686-708.

Hampton, J. A. (2006). Concepts as prototypes. In B. H. Ross (Ed.) *The Psychology of Learning and Motivation: Advances in research and theory* (pp. 79-113), 46. San Diego: Elsevier Academic Press.

Hampton, J. A. & Gardiner, M. M. (1983). Measures of internal category structure: A correlational analysis of normative data. *British Journal of Psychology*, 74, 491-516.

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S. & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Science of the United States of America*, 95, 15061-15065.

Hauk, O., Johnsrude, I. & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41, 301-307.

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

Hemeren, P. E. (1996). Frequency, ordinal position and semantic distance as measures of cross-cultural stability and hierarchies for action verbs. *Acta Psychologica*, 91, 39-66.

Hemeren, P. E. (1997). Typicality and context effects in action categories. In M. G. Shafto & P. Langley (Eds.) *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (p. 949). Hillsdale, NJ: Erlbaum.

Hemeren, P. E. (2003) Long-term priming and orientation specificity in the perception of biological motion. In T. Bajo & Lupiáñez (Eds.) *XIII Conference of the European Society of Cognitive Psychology* (p. 250). Universidad De Granada.

Hemeren, P. E. (2005). Orientation specific effects of automatic access to categorical information in biological motion perception. In B. G. Bara, L. W. Barsalou & M. Bucciarelli (Eds.) *Proceedings of the Twenty-seventh Annual Conference of the Cognitive Science Society* (pp. 935-940). Hillsdale, NJ: Erlbaum.

Hemeren, P. E. & Gawronska, B. (2007) Lexicalization of natural actions and cross-linguistic stability. In E. Ahlsén et al. (Eds.) *Communication - Action - Meaning: A Festschrift to Jens Allwood* (pp. 57-73). Göteborg University.

Hemeren, P.E., Kasviki, S. & Gawronska, B. (2008). Lexicalization of natural actions and cross-linguistic stability. In A. Botinis (Ed.) *Proceedings of the 2nd ISCA Workshop on Experimental Linguistics* (pp. 105-108). Athens, Greece: University of Athens.

Hemeren, P. E. & Pálsdóttir, S. (2003). Category effects and orientation specificity in inverted and upright displays of biological motion. *Perception*, 42, 142.

Hesslow, G. (2002). Conscious thought as simulation of behavior and perception. *Trends in Cognitive Sciences*, 6(6), 242-247.

Hill, H., Jinno, Y. & Johnston, A. (2003). Comparing solid-body with point-light animations. *Perception*, 32, 561-566.

Hill, H. & Pollick, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. *Psychological Science*, 11(3), 223-228.

Hirai, M. & Hiraki, K. (2006). The relative importance of spatial versus temporal structure in the perception of biological motion: An event-related potential study. *Cognition*, 99, B15-B29.

Hirai, M., Senju, A., Fukushima, H. & Hiraki, K. (2005). Active processing of biological motion perception: an ERP study. *Cognitive Brain Research, 23*, 387-396.

Hiris, E. (2007). Detection of biological and nonbiological motion. *Journal of Vision*, 7(12):4, 1-16.

Hiris, E., Krebeck, A., Edmonds, J. & Stout, A. (2005). What learning to see arbitrary motion tells us about biological motion perception. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 1096-1106.

Hochstein S. & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36, 791-804.

Hoenkamp, E. (1978). Perceptual cues that determine the labeling of human gait. *Journal of Human Movement Studies*, 4, 59-69.

Hoffman, D. D. & Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics*, 42, 195-204.

Hommel, B., Müsseler, J., Aschersleben, G. & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24, 849-937.

Hunt, A. R. & Halper, F. (2008). Disorganizing biological motion. *Journal of Vision*, 8(9), 12, 1-5, http://journalofvision.org/8/9/12/, doi:10.1167/8.9.12.

Huttenlocher, J. & Lui, F. (1979). The semantic organization of some simple nouns and verbs. *Journal of Verbal Learning and Verbal Behavior*, 18(2), 141-162.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C. & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PloS Biology*, 3(3), 0529-0535.

Ikeda, H., Blake, R. & Watanabe, K. (2005). Eccentric perception of biological motion is unscalably poor. *Vision Research*, 45, 1935-1943.

Jackendoff, R. (1987). On beyond zebra: The relation of linguistic and visual information. *Cognition*, 26, 89-114.

Jackendoff, R. (1990). *Semantic Structures*. Cambridge, MA: MIT Press.

Jackson P. L. & Decety, J. (2004). Motor cognition: a new paradigm to study self-other interactions. *Current Opinion in Neurobiology*, 14, 259-263.

Jacobs, A., Pinto, J. & Shiffrar, M. (2004). Experience, context, and the visual perception of human movement. *Journal of Experimental Psychology: Human Perception and Performance*, 30(5), 822-835.

Jacobs, A. & Shiffrar, M. (2005). Walking by perception by walking observers. *Journal of Experimental Psychology: Human Perception and Performance*, 31(1), 157-169.

Jastorff, J., Kourtzi, Z. & Giese, M. A. (2002). Learning of the discrimination of artificial complex biological motion. In R. P. Würz and M. Lappe (Eds.), *Dynamic Perception* (pp. 133-138), Infix, Berlin.

Jastorff, J., Kourtzi, Z. & Giese, M. A. (2006). Learning to discriminate complex movements: Biological versus artificial trajectories. *Journal of Vision*, 6, 791-804.

Jeannerod, M. (1994). The representing brain. Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17, 187-245.

Jeannerod, M. (2006). *Motor cognition: What actions tell the self.* Oxford: Oxford University Press.

Jeannerod, M. & Jacob, P. (2005). Visual cognition: a new look at the two-visual systems model. *Neuropsychologia*, 43, 301-312.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2), pp. 201-211.

Johansson, G. (1975). Visual Motion Perception. *Scientific American*, (June), pp 76-88.

Johnson, K. E. & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, 126(3), 248-277.

Jokisch, D., Daum, I., Suchan, B. & Troje, N. (2005). Structural encoding and recognition of biological motion: Evidence from event-related potentials and source analysis. *Behavioural Brain Research*, 157, 195-204.

Kalish, C. W. (1995). Essentialism and graded membership in animal and artifact categories. *Memory & Cognition*, 23(3), 335-353.

Klatzky, R. L., MacWhinney, B. & Behrmann, M. (Eds.) (2008). *Embodiment, Ego-Space, and Action*. New York: Psychology Press.

Klatzky, R. L., Pellegrino, J., McCloskey, B. P. & Lederman, S. J. (1993). Cognitive representations of functional interactions with objects. *Memory & Cognition*, 21 (3), 294-303.

Komatsu, L. K. (1992). Recent Views of Conceptual Structure. *Psychological Bulletin*, 112(3), 500-526.

Kourtzi, Z. & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience*, 12:1, 48-55.

Kourtzi, Z. & Shiffrar, M. (1999). Dynamic representations of human body movement. *Perception*, 28, 49-62.

Kozlowski, L. T. & Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics*, 21(6), 575-580.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. Chicago: The University of Chicago Press.

Lange, J. & Lappe, M. (2006). A model of biological motion perception from configural form cues. *The Journal of Neuroscience*, 26(11), 2894-2906.

Lange, J. & Lappe, M. (2007). The role of spatial and temporal information in biological motion perception. *Advances in Cognitive Psychology*, 3(4), 419-428.

Lange, J., Georg, K. & Lappe, M. (2006). Visual perception of biological motion by form: A template matching analysis. *Journal of Vision*, 6, 836-849.

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

Lestou, V., Pollick, F. E. & Kourtzi, Z. (2008). Neural substrates for action understanding at different description levels in the human brain. *Journal of Cognitive Neuroscience*, 20(2), 324-341.

Loula, F., Prasad, S. Harber, K. & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance*, 31(1), 210-220.

Lozano, S. C., Hard, B. M. & Tversky, B. (2007). Putting action in perspective. *Cognition*, 103, 480-490.

Lozano, S. C., Hard, B. M. & Tversky, B. (2008). Putting motor resonance in perspective. *Cognition*, 106, 1195-1220.

Ma, Y. L., Paterson, H. M. & Pollick, F. E. (2006). A motion-capture library for the study of identity, gender and emotion perception from biological motion. *Behavior Research Methods, Instruments, & Computers*. 38(1), 134-141.

Mahon, B. Z. & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology*, doi:1010.1016/j.physparis.2008.03.004.

Malt, B. C. (1995). Category Coherence in Cross-Cultural Perspective. *Cognitive Psychology*, 29, 85-148.

Malt, B. C., Gennari, S., Imai, M., Ameel, E., Tsuda, N. & Majid, A. (2008). Talking about walking: Biomechanics and the language of locomotion. *Psychological Science*, 19(3), 232-240.

Malt, B. C. & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory & Cognition*, Vol. 10 (1), 69-75.

Mandler, J. (2004). *The Foundations of Mind*. Oxford: Oxford University Press.

Margolis, E. & Laurence, S. (Eds.). (1999). *Concepts: Core Readings*. Cambridge, MA: MIT Press.

Marr, D. & Vaina, L. (1982). Representation and the recognition of the movement of shapes. *Proc. R. Soc. Lond.*, 24, 501-524.

Marr, D. & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. Lond.*, B 200, 1269-294.

Mather, G., Radford, K. & West, S. (1992). Low-level visual processing of biological motion. *Proceedings of the Royal Society of London, B249*, 149-155.

Mather, G. & West, S (1993). Recognition of animal locomotion from dynamic point-light displays. *Perception*, 22, 759-766.

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

McCloskey, M. & Glucksberg, S. (1978) Natural categories: Well-defined or fuzzy sets? *Memory & Cognition*, 6, 462-272.

McCloskey, M. & Glucksberg, S. (1979). Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology*, 11, 1-37.

McLeod, P., Dittrich, W., Driver, J., Perrett, D. & Zihl, J. (1996). Preserved and impaired detection of structure from motion by a “motion-blind” patient. *Visual Cognition*, 3(4), 363-391.

Medin, D. L. (1989). Concepts and Conceptual Structure. *American Psychologist*, Vol. 44, No. 12, 1469-1481.

Medin, D. L. & Barsalou, L. W. (1987). Categorization processes and categorical perception. In S. Harnad (Ed.) *Categorical Perception* (pp. 455-490). Cambridge: Cambridge University Press.

Medin, D. L., Lynch, E. & Solomon, K. O. (2000). Are there kinds of concepts?. *Annual Review of Psychology*, 51, 121-147.

Medin, D. L. & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.) *Similarity and Analogical Reasoning* (pp.179-195). New York: Cambridge Univ. Press.

Medin, D. L., & Smith, E. E. (1984). Concepts and Concept Formation. *Annual Review of Psychology*, 35, 113-138.

Meltzoff, A. N. & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198, 75-78.

Meltzoff, A. N. & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, 54, 702-709.

Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.) *Concepts and Conceptual Development* (pp. 201-233). Cambridge: Cambridge University Press.

Mervis, C. B., Catlin, J. & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, 7, 283-284.

Mervis, C. B. & Crisafi, M. A. (1982). Order of Acquisition of Subordinate-, Basic-, and Superordinate-Level Categories. *Child Development*, Vol. 53, 258-266.

Mervis, C. B. & Rosch, E. (1981). Categorization of Natural Objects. *Annual Review of Psychology*, 32, 89-115.

Miller, G. A. & Johnson-Laird, P. N. (1976). *Language and Perception*. Cambridge University Press.

Morris, M. W. & Murphy, G. L. (1990). Converging operations on a basic level in event taxonomies. *Memory & Cognition*, 18 (4), 407-418.

Murphy, G. L. (1991a). Parts in object concepts: Experiments with artificial categories. *Memory & Cognition*, 19(5), 423-438.

Murphy, G. L. (1991b). More on parts in object concepts: Response to Tversky and Hemenway. *Memory & Cognition*, 19(5), 443-447.

Murphy, G. L. (2002). *The Big Book of Concepts*. Bradford Books, MIT Press.

Murphy, G. L. (2003). The downside of categories. *Trends in Cognitive Sciences*, 7(12), 513-514.

Murphy, G. L. & Brownell, H. H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning Memory, and Cognition*, 11(1), 70-84.

Murphy, G. L. & Ross, B. H. (2005). The two faces of typicality in category-based induction. *Cognition*, 95, 175-200.

Murphy, G. L. & Smith, E. E. (1982). Basic-Level Superiority in Picture Categorization. *Journal of Verbal Learning and Verbal Behavior*, Vol. 21, 1-20.

Neisser, U. (1987). From direct perception to conceptual structure. In U. Neisser (Ed.) *Concepts and Conceptual Development*. Cambridge: Cambridge University Press.

Neri, P., Luu, J. Y. & Levi, D. M. (2007). Sensitivity to biological motion drops by ~1/2 log-unit with inversion, and is unaffected by amblyopia. *Vision Research*, 47, 1209-1214.

Newtson, D. (1973). Attribution and the Unit of Perception of Ongoing Behavior. *Journal of Personality and Social Psychology*, 28(1), 28-38.

Newtson, D. & Engquist, G. (1976). The Perceptual Organization of Ongoing Behavior. *Journal of Experimental Social Psychology*, 12, 436-450.

Newtson, D., Engquist, G. & Bois, J. (1977). The Objective Basis of Behavior Units. *Journal of Personality and Social Psychology*, 35(12), 847-862.

Nijsse, M. (1988). Testing the Significance of Kendall's τ and Spearman's r_S. *Psychological Bulletin,* 103(2), 235–237.

Olofsson, U., Nyberg, L. & Nilsson, L-G. (1997). Priming and recognition of human motion patterns. *Visual Cognition*, 4, 373-382.

Palmeri, T. J. & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, 5, 291-303.

Pavlova, M., Lutzenberger, W., Sokolov, A. & Birbaumer, N. (2004). Dissociable cortical processing of recognizable and non-recognizable biological movement: Analysing gamma MEG activity. *Cerebral Cortex*, 14, 181-188.

Pavlova, M. & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception & Psychophysics*, 62(5), 889-899.

Pavlova, M. & Sokolov, A. (2003). Prior knowledge about display inversion in biological motion perception. *Perception*, 32, 937-946.

Peelen, M. V., Wiggett, A. J. & Downing, P. E. (2006). Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron*, 49, 815-822.

Peuskens, H., Vanrie, J., Verfaillie, K. & Orban, G. A. (2005). Specificity of regions processing biological motion. *European Journal of Neuroscience*, 21, 2864-2875.

Pinto, J. (2006). Developing body representations: A review of infants' responses to biological-motion displays. In G. Knoblich, I. M. Thornton, M. Grosjean & M. Shiffrar (Eds.) *Human body perception from the inside out* (pp. 305-322). Oxford: Oxford Univ. Press.

Pinto, J. & Shiffrar, M. (1999). Subconfigurations of the human form in the perception of biological motion displays. *Acta Psychologica*, 102, 293-318.

Pollick, F. E. (2004). The features people use to recognize human movement style. In A. Camurri & G. Volpe (Eds.) *Gesture-Based Communication in Human-Computer Interaction* (Vol. 2915, pp. 10-19). Berlin: Springer-Verlag Berlin.

Pollick, F. E., Fidopiastis, C. & Braden, V. (2001). Recognising the style of spatially exaggerated tennis serves. *Perception*, 30, 323-338.

Pollick, F. E., & Kourtzi, Z. (1998). The visual perception of human arm movements. *Investigative Opthamology & Visual Science*, 39(4), S1094.

Prinz, W. (2006). What re-enactment earns us. *Cortex*, 42, 515-517.

Prinz, W. & Hommel, B. (Eds.) (2002). *Common mechanisms in perception and action: Attention and Performance*, 19. Oxford: Oxford University Press.

Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Bradford Books, MIT Press.

Puce, A. & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Phil. Trans. R. Soc. Lond. B*, 358, 435-445.

Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, 5(12), 517-524.

Pulvermüller, F., Härle, M. & Hummel, F. (2001). Walking or talking: Behavioral and neurophysiological correlates of action verb processing. *Brain and Language*, 78, 143-168.

Pyles, J. A., Garcia, J. O., Hoffman, D. D. & Grossman, E. D. (2007). Visual perception and neural correlates of novel 'biological motion.' *Vision Research*, 47, 2786-2797.

Reed, C. L., Stone, V. E., Bozova, S. & Tanaka, J. (2003). The body-inversion effect. *Psychological Science*, 14(4), 302-308.

Reed, C. L., Stone, V. E. & McGoldrick, J. E. (2006). Not just posturing: Configural processing of the human body. In G. Knoblich, I. M. Thornton, M. Grosjean & M. Shiffrar (Eds.) *Human body perception from the inside out* (pp. 229-258). Oxford: Oxford Univ. Press.

Reed, C. L., Stone, V. E., Grubb, J. D. & McGoldrick, J. E. (2006). Turning configural processing upside down: Part and whole body postures. *Journal of Experimental Psychology: Human Perception and Performance*, 32(1), 73-87.

Rifkin, A. (1985). Evidence for a basic level in event taxonomies. *Memory & Cognition*, 13, 538-556.

Rips, L. J., Shoben, E. J. & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 12, 1-20.

Rizzolatti, G. (2004). Understanding the actions of others. In N. Kanwisher & J. Duncan (Eds.) *Functional Neuroimaging of Visual Cognition: Attention and Performance XX* (pp. 3-26). New York: Oxford University Press.

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

Rizzolatti, G., Fogassi, L. & Gallese, V. (2006). Mirrors in the mind. *Scientific American*, November, 54-61.

Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3, 131-141.

Rizzaolatti, G., Fogassi, L. & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2, 661-670.

Rosch, E. (1973). Natural Categories. *Cognitive Psychology*, 4, 328-350.

Rosch, E. (1975). Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General*, 104(3), 192-233.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.) *Cognition and Categorization* (pp.27-48). Hillsdale, NJ: Lawrence Erlbaum.

Rosch, E. & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.

Rosch, E. & Mervis, C. B. (1981). Categorization of natural objects. *Annual Review of Psychology*, 32, 89-115.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. (1976). Basic Objects in Natural Categories. *Cognitive Psychology*, 8, 382-439.

Rosch, E., Simpson, C. & Miller, R. S. (1976). Structural Bases of Typicality Effects. *Journal of experimental Psychology: Human Perception and Performance*, 2(4), 491-502.

Ross, B. H. & Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology*, 38, 495-553.

Runeson, S. & Frykholm, G. (1981). Visual Perception of Lifted Weight. *Journal of Experimental Psychology: Human Perception and Performance*, 7(4), 733-740.

Runeson, S. & Frykholm, G. (1983). Kinematic Specification of Dynamics as an Informational Basis for Person-and-Action Perception: Expectation, Gender Recognition, and Deceptive Intention. *Journal of Experimental Psychology: General*, 112(4), 585-615.

Saygin, A. P., Wilson, S. M., Hagler, D. J., Bates, E. & Sereno, M. I. (2004a). Point-light biological motion perception activates human premotor cortex. *Journal of Neuroscience*, 24(27), 6181-6188.

Saxe, R., Xiao, D.-K., Kovacs, G., Perrett, D. I. & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, 42, 1435-1446.

Schank, R. C. & Abelson, R. P. (1977). *Scripts, Plans, Goals, and Understanding*. Hillsdale, NJ: Erlbaum.

Schyns, P. G. (1997). Categories and percepts: a bi-directional framework for categorization. *Trends in Cognitive Sciences*, 1(5), 183-189.

Sebanz, N., Knoblich, G. & Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition*, 88, B11-B21.

Shapiro, K. A., Moo, L. R. & Caramazza, A. (2006). Cortical signatures of noun and verb production. *Proceedings of The National Academy of Sciences of the USA*, 103(5), 1644-1649.

Shepard, R.N. (1962a). The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function. I. *Psychometrika*, Vol. 27, No. 2, 125–140.

Shepard, R.N. (1962b). The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function. II. *Psychometrika*, Vol. 27, No. 3, 219–246.

Shiffrar, M. (2006). Body-based views of the world. In G. Knoblich, I. M. Thornton, M. Grosjean & M. Shiffrar (Eds.) *Human body perception from the inside out* (pp. 135-146). Oxford: Oxford Univ. Press.

Shiffrar, M. (2008). Embodied motion perception: Psychophysical studies of the factors defining visual sensitivity to self- and other-generated actions. In R. L. Klatzky, B. MacWhinney & M. Behrmann (Eds.) *Embodiment, Ego-Space, and Action* (pp. 113-143). New York: Psychology Press.

Shiffrar, M. & Freyd, J. J. (1993). Timing and apparent motion path choice with human body photographs. *Psychological Science*, 4, 379-384.

Shiffrar, M., Lichtey, L. & Heptulla Chatterjee, S. (1997). The perception of biological motion across apertures. *Perception and Psychophysics*, 59(1), 51-59.

Shiffrar, M. & Pinto, J. (2002). The visual analysis of bodily motion. In W. Prinz & B. Hommel (Eds.) *Common mechanisms in perception and action: Attention and Performance*, 19 (pp. 381-399). Oxford: Oxford University Press.

Shipley, T. F. (2003). The effect of object and event orientation on perception of biological motion. *Psychological Science*, 14(4), 377-380.

Shipley, T. F. & Brumberg, J. S. (2005). Markerless motion-capture for point-light displays. (Tech. Rep.). Temple University, Department of Psychology. http://astro.temple.edu/~shipley/mocap/MarkerlessMoCap.pdf. [accessed: 2006-04-18].

Shoben, E. (1982). Semantic and lexical decisions. In C. R. Puff (Ed.) *Handbook of research methods in human memory and cognition* (pp. 287-314). New York: Academic Press.

Slobin, D. I. (2004). The many ways to search fro a frog: Linguistic typology and the expression of motion events. In S. Strömqvist & L. Verhoeven (Eds.) *Relating events in narrative: Vol. 2. Typological and contextual perspectives* (pp. 219-257). Mahwah, NJ: Lawrence Erlbaum Associates.

Smith, E. E. & Medin, D. L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Solomon, K. O., Medin, D. L. & Lynch, E. (1999). Concepts do more than categorize. *Trends in Cognitive Sciences*, 3(3), 99-105.

Stevens, S. S. (1951). Mathematics, Measurement and Psychophysics. In S. S. Stevens (Ed.) *Handbook of Experimental Psychology* (pp. 1-49). New York: Wiley.

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

Svensson, H., Lindblom, J. & Ziemke, T. (2007). Making sense of embodied cognition: Simulation theories of shared neural mechanisms for sensorimotor and cognitive processes. In T. Ziemke, R. M. Frank & J. Zlatev (Eds.) *Body, Mind and Language. Volume 1: Embodiment* (pp. 241-270). Berlin: Mouton de Gruyter.

Tai, Y. F., Scherfler, C., Brooks, D. J., Sawamoto, N. & Castiello, U. (2004). The human premotor cortex is 'Mirror' only for biological actions. *Current Biology*, 14, 117-120.

Tanaka, J. W. & Taylor, M. E. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, 15, 121-149.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P. et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, 17(2), 273-281.

Thomas, S. M. & Jordan, T. R. (2001). Techniques for the production of point-light and fully illuminated video displays from identical recordings. *Behavior Research Methods, Instruments, & Computers*, 33(1), 59-64.

Thompson, B., Troje, N. F., Hansen, B. C. & Hess, R. F. (2008). Amblyopic perception of biological motion. Journal of Vision, 8(4):22, 1-14, http://journalofvision.org/8/4/22/, doi:10.1167/8.4.22.

Thornton, I. M. (2006). Biological motion: Point-light walkers and beyond. In G. Knoblich, I. M. Thornton, M. Grosjean & M. Shiffrar (Eds.) *Human body perception from the inside out* (pp. 271-303). Oxford: Oxford Univ. Press.

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

Thornton, I. M., Rensink, R. A. & Shiffrar, M. (2002). Active versus passive processing of biological motion. *Perception*, 31, 837-853.

Thornton, I. M., Vuong, Q. C. & Bülthoff, H. H. (2003). A chimeric point-light walker. *Perception*, 32, 377-383.

Thornton, I. M. & Vuong, Q. C. (2004). Incidental processing of biological motion. *Current Biology*, 14, 1084-1089.

Thurman, S. M. & Grossman, E. D. (2008). Temporal "Bubbles" reveal key features for point-light biological motion perception. Journal of Vision, 8(3):28, 1-11, http://journalofvision.org/8/3/28/, doi:10.1167/8.3.28.

Tomasino, B., Fink, G. R., Sparing, R., Dafotakis, M. & Weiss, P. H. (2008). Action verbs and the primary motor cortex: A comparative TMS study of silent reading, frequency judgments, and motor imagery. *Neuropsychologia*, doi:101016/j.neuropsychologia.
2008.01.015.

Tomasino, B., Werner, C. J., Weiss, P. H. & Fink, G. R. (2007). Stimulus properties matter more than perspective: An fMRI study of mental imagery and silent reading of action phrases. *NeuroImage*, 36, T128-T141.

Tranel, D., Kemmerer, D., Adophs, R., Damasio, H. & Damasio, A. R. (2003). Neural correlates of conceptual knowledge for actions. *Cognitive Neuropsychology*, 20, 409-432.

Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.

Troje, N. F. (2003). Reference frames for orientation anisotropies in face recognition and biological-motion perception. *Perception*, 32, 201-210.

Troje, N. F. & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a "life detector"? *Current Biology*, 16, 821-824.

Troje, N. F., Westhoff, C. & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematics cues. *Perception & Psychophysics*, 67(4), 667-675.

Turella, L., Pierno, A. C., Rubaldi, F. & Castiello, U. (2008). Mirror neurons in humans: Consisting or confounding evidence? *Brain and Language*, doi:10.1016/j.bandl.2007.11.002

Tversky, B. & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, 113(2), 169-193.

Tversky, B. & Hemenway, K. (1991). Parts and the basic level in natural categories and artificial stimuli: Comments on Murphy (1991). *Memory & Cognition*, 19(5), 439-442.

Vaina, L. M. & Gross, C. G. (2004). Perceptual deficits in patients with impaired recognition of biological motion after temporal lobe lesions. *Proceedings of the National Academy of Science of the United States of America*, 101(48), 16947-16951.

Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Science of the United States of America*, 98(20), 11656-11661.

Vallacher, R. R., & Wegner, D. M. (1987). What Do People Think They're Doing? Action Identification and Human Behavior. *Psychological Review*, 94(1), 3-15.

Vanrie, J., Dekeyser, M. & Verfaillie, K. (2004). Bistability and biasing effects in the perception of ambiguous point-ligth walkers. *Perception*, 33, 547-560.

Vanrie, J. & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers*, 36, 625-629.

VanRullen, R. & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception*, 30, 655-668.

Verfaillie, K. (1993). Orientation-dependent priming effects in the perception of biological motion. *Journal of Experimental Psychology: Human Perception & Performance*, 19, 992-1013.

Verfaillie, K, (2000). Perceiving human locomotion: Priming effects in direction discrimination. *Brain and Cognition*, 44, 192-213.

Verfaillie, K. & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition*, 9(1/2), 217-232.

Verfaillie, K., De Troy, A. & Van Rensbergen, J. (1994). Transsaccadic integration of biological motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(3), 649-670.

Vetenskapsrådet (2002). *Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning*. Vetenskapsrådet, Sverige.

Viberg, Å. (1992). Universellt och språkspecifikt i det svenska ordförrådets organisation. *Tijdschrift voor Skandinavistiek*, 13:2, 17–58.

Vigliocco, G., Vinson, D. P., Lewis, W. & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, 48, 422-488.

Vinson, D. P. & Vigliocco, G. (2002). A semantic analysis of grammatical class impairments: semantic representations of object nouns, action nouns and action verbs. *Journal of Neurolinguistics*, 15, 317-351.

Viviani, P. (2002). Motor competence in the perception of dynamic events: a tutorial. In W. Prinz & B. Hommel (Eds.) *Common mechanisms in perception and action: Attention and Performance*, 19 (pp. 406-442). Oxford: Oxford University Press.

von Linné, C. (1756/1962). *En Gammal Svensk hundbok*. Sigtuna: Bokförlaget Fabel.

Vuong, Q. C., Hof, A. F., Bülthoff, H. H. & Thornton, I. M. (2006). An advantage for detecting dynamic targets in natural scenes. *Journal of Vision*, 6(1), 87-96.

Webb, J. A. & Aggarwal, J. K. (1982). Structure from motion of rigid and jointed objects. *Artificial Intelligence*, 19, 107-130.

Wilson, M. (2006). Covert imitation: How the body schema acts as a prediction device. In G. Knoblich, I. M. Thornton, M. Grosjean & M. Shiffrar (Eds.) *Human body perception from the inside out* (pp. 211-228). Oxford Univ. Press, Oxford, 211-228.

Wolfe, J. M. & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5(6), 495-501.

Yeh, W. & Barsalou, L. W. (2006). The situated nature of concepts. *American Journal of Psychology*, 119(3), 349-384.

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

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

Ziemke, T., Frank, R. M. & Zlatev, J. (Eds.) (2007). *Body, Mind and Language. Volume 1: Embodiment* (pp. 241-270). Berlin: Mouton de Gruyter.

Zwan, R. A. & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, 135(1), 1-11.