PREDICTIVE GAZE IN ACTION OBSERVATION:
SOCIAL LEARNING IN ACTION

by

Andreas Reichelt

A thesis submitted to the Centre for Neuroscience Studies
In conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
(September, 2015)

Copyright © Andreas Reichelt, 2015
Abstract

How do we observe other people engaged in activities? Predictive gaze in action observation was first described in 2003 by Flanagan and Johansson who showed that in a block stacking task, observers, like actors, shifted their gaze to locations of upcoming contact events (objects to be grasped and object placement sites) around the same time that a hand movement was initiated. Later studies have shown that even when observers do not know what the actor is going to do next ahead of time, their gaze is still robustly drawn to contact events, typically arriving there before the actor's hand based on extrapolation of the ongoing action.

Here I introduce an organizational framework for gaze behavior in action observation in terms of distinct modes of predictive gaze – anticipation, extrapolation, and tracking – and contextual factors shaping observer gaze behavior. These factors include the scene configuration as well as the timing, kinematics, and goals of the actor's movements, but also crucially depend on observers, their knowledge and skills, their perspective on the scene, as well as their own interest and goals. In the three studies presented here we investigate the proposal that observers, like actors, seek to closely monitor object contact events, including object lift-off, in part to learn about and keep track of object properties in the service of guiding future actions. In chapter 2 I describe and quantify social motor learning of object weight. In chapter 3 I show that when observing a demonstrator lift two objects at the same time, participants preferentially shift their gaze towards objects which they expect to subsequently act on themselves. In chapter 4 I describe how knowing the circumstances of an action (object value and distance) translates into predictive gaze behavior in action observation, characterize distinct gaze strategies, and evaluate their visual consequences for the observer. In the final chapter I apply this framework to critically review the current literature – which tends to conflate modes of predictive gaze – revisit the relationship between observer gaze behavior and the mirror neuron system, and review studies into the developmental trajectory of predictive gaze in action observation.
Co-Authorship

The research in all the three studies presented in this thesis has been conducted under the supervision of Randy Flanagan. In each study we have been working with undergraduate students, Alyssa Ash, Natasha Bowman, and Percy Chan, respectively. Roland Johansson has contributed to analysis and discussion of the result of the study presented in chapter 2. I was involved in all aspects of the work and am listed as primary author for all three studies.

Chapter 2

Chapter 3

Chapter 4

This work is licensed under a Creative Commons Attribution 4.0 International License (Attribution license - CC BY)
Acknowledgements

I have been fortunate to have crossed paths, conversed, and collaborated with many interesting people during my five years at the Centre for Neuroscience Studies, both within and outside of academia. I would like to acknowledge first and foremost the unwavering support of my supervisor Randy Flanagan whose sense of wonder and adventure at how we move small cubes around and his knack for doing science at any and all times shall always be an inspiration. I also want to thank Lee Baugh for his calm and curious spirit, as well as my fellow graduate students who taught me some of the ropes – Brandie Stewart, Jonathan Diamond, and Simona Jantz – which I in turn have done my best to pass on to Ashley Bramwell, Josh Moskowitz, and Tiger Ji. Thank you all for not curbing your enthusiasm ;­). This still leaves the other post-docs in the lab – Aarlenne Khan, then Kevin Trewartha and Jason Gallivan, as well as Fred Danion our collaborator. I’d like to particularly mention the honor students and research assistants I had the pleasure of collaborating with over the years: Alyssa Ash, James Kim, Percy Chan, and Natasha Bowman. Very special nod to Kelly Woudsma for all the fun we had breaking the boundaries of the cold steel of industrialized science. Speaking of which, it has been a particular pleasure to watch Sean Hickman work magic with metal, wood, and synthetics – really any material other than fabric which clearly cannot be trusted. And last but not least, I shall miss the company, C code, and electronics wizardry of Martin York who taught me how to coil cables properly.

Beyond our lab, I'd like to thank everyone on my committee for their support, in particular John Kalaska for his detailed notes which were both meticulous and humorous at the same time, and to Steve Scott and Val Kuhlmeier for engaging with the material “from below” (motor control) and “from above” (developmental psychology). Also thanks to everyone from the CNS – particularly Ethan Heming who has been my landlord, colleague, and overall too good a person even by Canadian standards, Andrew Pruszynski, Ashley Parr, Benedict Chang, Brian Coe, Carl Jackson, Dominic Standage, Fred Crevecoeur, Gaurav Aggarwal, Helen Bretzke, Janis Kan, Jay Jantz, Justin Peterson, Mike Lewis, Mohsen Omrani, Parisa Khoozani, Rachel Bosma and Teige Bourke, Saba Farbodkia, Scott Murdison, Sisi Xu, and all the rest of you guys. And of course thanks to Scott Robson and Kathleen Merwin for your efforts to carefully attach home made electronics to infant's heads with Nicole and
me. Turning from studying action to activism, I shall fondly remember our tilling against windmills as Queen's resident freethinkers, OcQpiers, and bitumen zombies, working with OPIRG and the SGPS to keep our spirits up in a confused world.

A big thank you to my family, in particular my sister and mother for not letting me go even after I crossed a large body of water, and to my grandmother who wrapped up her centennial while I ran my extra lap. To my partner Nicole who never left my side or my thoughts while I had gone far away to lift small objects. Finally, this thesis is dedicated to the memory of my father.
# Table of Contents

Abstract........................................................................................................................................................................... i

Co-Authorship........................................................................................................................................................................... ii

Acknowledgements....................................................................................................................................................................... iii

List of Figures............................................................................................................................................................................... ix

List of Tables................................................................................................................................................................................... xi

List of Abbreviations...................................................................................................................................................................... xii

Chapter 1 General introduction....................................................................................................................................................... 1

1.1 Action Observation in Naturalistic Interactions.................................................................................................................. 1

1.2 Regularities in gaze patterns and their sources...................................................................................................................... 3

1.2.1 Bottom-up and top-down approaches to visual saliency................................................................................................. 3

1.2.2 Eye movements in naturalistic activities and eye-hand coordination............................................................................. 6

1.2.3 Observer-Actor eye movement coupling in action observation....................................................................................... 7

1.3 Action Plans in Action Observation Revisited.................................................................................................................... 9

1.3.1 Situating block stacking....................................................................................................................................................... 9

1.4 An organizational framework of predictive gaze in action observation........................................................................... 12

1.4.1 Modes of predictive gaze.................................................................................................................................................... 12

1.4.2 A framework of factors determining modes of predictive gaze....................................................................................... 14

1.4.3 Prior Knowledge of Actions, Scene and Movement Metrics, and Observer Goals as Determinants of Gaze Behavior in Action Observation........................................................................................................... 17

1.5 From Action Observation to Social Learning in Action...................................................................................................... 21

1.5.1 Chapter 2........................................................................................................................................................................... 23

1.5.2 Chapter 3........................................................................................................................................................................... 23

1.5.3 Chapter 4........................................................................................................................................................................... 24

1.6 References................................................................................................................................................................................. 25
5.3 The relationship between predictive gaze behavior in action observation and the mirror neuron system. 140

5.3.1 Goal prediction, action understanding, and learning from observing .................................................. 140

5.3.2 The mirror neuron system and predictive gaze based on extrapolation ........................................... 142

5.3.3 The mirror neuron system and anticipatory predictive gaze .......................................................... 145

5.4 Developing a capacity and inclination for predictive gaze behavior ....................................................... 147

5.4.1 Expertise and practice effects in action observation ........................................................................ 147

5.4.2 Motor abilities as prerequisites for action observation: 3 caveats ..................................................... 148

5.4.3 Developmental trajectories for engaging in predictive gaze in action observation ......................... 150

5.4.4 Conclusions ........................................................................................................................................ 156

5.5 Conclusion and Outlook ....................................................................................................................... 156

5.5.1 Gaze behavior as interaction ........................................................................................................... 156

5.5.2 Social Learning in Action between Motor Control and Social Neuroscience ................................... 157

5.6 References ............................................................................................................................................ 158
List of Figures

Figure 1.1 Example clip from a naturalistic group activity.................................................................2
Figure 1.2 Architecture for computing saliency maps........................................................................4
Figure 1.3 Scanpaths from Yarbus (1967)........................................................................................5
Figure 1.4 Eye–hand coordination in action and action observation..................................................8
Figure 1.5 LEGO Train building exploration.....................................................................................11
Figure 1.6 Modes of predictive gaze..................................................................................................13
Figure 1.7 Design of Rotman et al. (2006)........................................................................................18
Figure 2.1 Experimental set-up and data analysis.............................................................................34
Figure 2.2 Lifting traces from a single participants..........................................................................40
Figure 2.3 Distributions of initial peak LF rates and maximum lift heights........................................42
Figure 2.4 Average LF results............................................................................................................44
Figure 2.5 Average GF results............................................................................................................47
Figure 2.6 Average lift height results..................................................................................................50
Figure 3.1 Experimental setup...........................................................................................................65
Figure 3.2 Individual results participant P1.......................................................................................73
Figure 3.3 Individual results participant P7.......................................................................................75
Figure 3.4 Individual results participant P8.......................................................................................76
Figure 3.5 Group results experiment 1...............................................................................................79
Figure 3.6 Demonstrator lifts.............................................................................................................82
Figure 3.7 Group results experiment 2...............................................................................................83
Figure 4.1 Experimental setup...........................................................................................................94
Figure 4.2 Actor choice behavior......................................................................................................99
Figure 4.3 Derivation of the criterion for proactive gaze initiation...................................................102
Figure 4.4 Individual results............................................................................................................104
Figure 4.5 Group results ................................................................. 107
Figure 4.6 Effects of target configuration ........................................ 109
Figure 4.7 Gaze strategies and their visual consequences ............... 111
Figure 5.1 Main and supplemental results from Falck-Ytter et al. (2006) 128
Figure 5.2 Analysis of predictive gaze in analogy to reaction time .......... 138
Figure 5.3 Stimulus material from Ambrosini et al. .............................. 143
List of Tables

Table 1.1 Framework of factors determining modes of predictive gaze behavior..............................15
Table 3.1: Order of conditions and trial composition of experiment 1.........................................................68
Table 3.2: Order of conditions and trial composition of experiment 2.........................................................69
Table 4.1: Graphical illustration of the order of experimental conditions.................................................96
Table 5.1: Follow up studies to Flanagan & Johansson (2003).................................................................131
Table 5.2: Studies reporting effects of contextual factors........................................................................132
Table 5.3: Selected studies on the development of predictive gaze in action observation......................150
List of Abbreviations

9m 9 months
ANOVA analysis of variance
EEG electroencephalogram
EOG electrooculogram
F5 ventral premotor cortex (PMv) according to the convention by Matelli et al. (1985)
FEF frontal eye field
LoS Line of Sight
M1 primary motor cortex
MNS mirror neuron system
PMv ventral premotor cortex
pSTS posterior part of the superior temporal sulcus
RT reaction time
rTMS repetitive transcranial magnetic stimulation
SEM standard error of the mean
TMS transcranial magnetic stimulation
vF vertical force
vF Rate rate of change of vertical force
Chapter 1

General introduction

1.1 Action Observation in Naturalistic Interactions

As social animals we constantly encounter other people and spend a lot of our time communicating with them, engaging in shared activities, or simply observing their activities. In this thesis I investigate gaze behavior in action observation as a basic form of social coordination. Action observation may occur over long durations and hence merits study in its own right, and interactive activities may include episodes of action observation interspersed with periods of joint action. How do we observe other people engaging in activities and how does this change our own engagement with the world?

Starting in the 1990s, the group around Michael Land has pioneered eye-tracking studies of real life cultural activities, such as driving (Land & Lee, 1994), tea making (Land, Mennie, & Rusted, 1999), and cricket (Land & McLeod, 2000) using a light-weight head mounted eye tracker. However, their series of studies (Land & Tatler, 2009) and studies inspired by them (Land, 2006; Hayhoe & Rothkopf, 2011) have so far been documenting gaze behavior in cultural activities that focus on individual rather than social actions. While researchers of visual perception have begun to include social content in the course of turning to more naturalistic stimulus statistics in scene processing (e.g. Võ, Smith, Mital, & Henderson, 2012), their studies typically focus on specific perceptual concerns and use a design based on participants passively watching video clips. However, only a few sporadic attempts have been made to validate laboratory stimulus materials (Dorr, Martinetz, Gegenfurtner, & Barth, 2010; Foulsham, Walker, & Kingstone, 2011), and the lack of interactivity of video clips may potentially change gaze allocation drastically (Dicks, Button, & Davids, 2010; Laidlaw, Foulsham, Kuhn, & Kingstone, 2011).
Notwithstanding the lack of controlled studies on action observation in naturalistic social interactions, it does seem quite clear that, generally speaking, actions of people robustly drive gaze in a scene. This is illustrated in a vignette taken from an exploratory study of gaze behavior in a joint cooking activity involving a large number of people that I conducted together with Nicole Rossmanith (unpublished observations). An outtake from the video transcript of 10 second duration is shown in figure 1.1.

![Figure 1](image)

**Figure 1** Example clip from a naturalistic group activity as documented by a participant wearing a lightweight eye tracker (pupil labs, open source technology see Kassner, Patera, & Bulling, 2014). Top: 5 still images (1 every 2 seconds) from the head camera. Pink dot shows gaze fixations events which are connected with solid pink lines if less than 500ms apart. Below we show the transcript of the activity using ELAN video annotation (free software, Max-Planck-Institut for Psycholinguistics, Nijmegen, Netherlands). In this 10 second clip gaze is directed towards knife cutting actions of fruit, especially to a 3 second long action where peach slices are cut and dropped into a bowl. Then gaze is shifted towards the toddler on the left, and his picking up of a mango.

Clearly the observer encounters a complicated buzz of activity with multiple candidate locations for gaze at a given time. Still, object oriented actions predominantly draw gaze whereas faces are looked at only intermittently unless the observer is engaged in face to face conversation. Why are observers so

---

1 There is a sizable literature analyzing video recordings of (semi-)naturalistic social interactions including manually coding of gaze drawing on approaches from ethnography and ethology (e.g. Goodwin, 1994, 2013; de Barbaro, Johnson, Forster, & Deak, 2013).
intensely interested in action observation? To address this question, I will first briefly review general accounts of what drives gaze allocation as a background before introducing our own approach based on action monitoring and control.

1.2 Regularities in gaze patterns and their sources

1.2.1 Bottom-up and top-down approaches to visual saliency

What drives eye-movements – where and when people direct their gaze and for how long – in natural gaze behavior? According to an influential computational neuroscience approach, the current visual stimulus array itself primarily determines how informative (salient) each location in the scene is. To compute a saliency map from an input scene (or video clip), bitmaps are analyzed in terms of parallel processing streams dedicated to color, intensity, orientation, movement, etc., and the resulting contrast maps are normalized and integrated into an overall saliency map used to predict the likelihoods of upcoming eye movement(s) of a generic observer (Itti, Koch, & Niebur, 1998; Itti & Koch, 2001, see figure 1.2).
However, the potentially decisive influence of “top-down” aspects have been documented in great detail already in the early days of eye-tracking. In a seminal study from the 1950s led by Alfred Yarbus, observers were shown an artwork (Repin's *The Unexpected Visitor*) with observer instructions varied by condition. Yarbus showed that scanpaths were systematically altered depending on the task at hand as participants were visually exploring regions of the painting which are intuitively relevant to the high-level questions posed (Yarbus, 1967, see figure 1.3). The influence of task instructions on gaze strategies has been essentially replicated with more modern (and now painless) eye tracking methods, with the potential

Figure 2 Architecture for computing saliency maps based on low-level image features processed in parallel. High contrast locations are identified within each channel and combined into an overall saliency map. Top-down factors are integrated as additional parameters for weighing the contributions of each feature channel. From Itti & Koch (2001).
caveat of much shorter overall exploration durations in self-paced conditions (i.e. DeAngelus & Pelz, 2009).

Figure 3: Scanpaths of one participant visually exploring the Repin’s Unexpected Visitor for several minutes in 4 selected conditions. It is instructive to first look at the pictures and attempt to guess the instructions working backwards from the scanpath data before reading on. Instructions given for the 4 example conditions are translated as: a free viewing/no instructions; b “identify the ages of the people”; c “estimate the material circumstances of the family”; and d “estimate how long the unexpected visitor has been away”. Reconstructed from scanpath data from Yarbus (1967).

Bottom-up computational models have been extended to address the influence of aspects particular to an observer – their prior knowledge and the information they seek to extract from the scene. To this end, observer concerns must somehow be coded on a conceptual level and integrated with the logic of machine vision. This is typically based on computer vision algorithms for identifying objects in a scene (“chair”) and parsing them into object classes (“furniture”). Instructions need to be issued in simple terms according to this level of scene “understanding” (e.g. “look at people, Navalpakkam & Itti, 2005) – and do not reach anywhere near the level of complexity inherent in Yarbus' task instructions (“estimate how long the unexpected visitor has been away”). Tellingly, studies guided by this “scene grammar” approach were initially unsuccessful at “replicating Yarbus”, i.e. such simplified task instructions did not reliably affect summary statistics of observer gaze behavior (Greene, Liu, & Wolfe, 2012). Partial predictive power was finally achieved after the task was made somewhat more meaningful to observers and more detailed statistical methods were used (Borji & Itti, 2014; see also the discussion in Haji-Abolhassani & Clark, 2014). Accounting for selective attention in complex naturalistic situations in greater detail would
seem to entail a much deeper appreciation of the interests and goals of individual observers and presents a challenge to an approach based purely on the current computer vision approach (see Tatler, 2009). This serves as a potent reminder that people can be expected to pursue gaze strategies particular to their individual quest for making sense of the world. Thus participants in a given experimental situation may show highly diverse gaze behaviors which are meaningful but may not readily be linked to experimental variables.

1.2.2 Eye movements in naturalistic activities and eye-hand coordination

Finally, when people are engaged in activities themselves, their gaze strategies change fundamentally with gaze typically becoming closely tied to their actions in time and space. In an influential series of studies using a lightweight head-mounted eye tracker, Land and colleagues have documented gaze behavior in everyday activities such as making tea (see above). While describing qualitative differences between activities, they found that within a given activity “most eye movements are closely and purposefully linked to the ongoing actions” (Land et al., 1999; Land & Tatler, 2009). Land developed a framework classifying functions of eye movement in action, including searching for objects, directing object-oriented actions, guiding object actions, and monitoring activity progress (Land, 2006). This observed link between gaze and action control has been confirmed and described in more detail in a controlled laboratory study investigating eye hand coordination in object manipulation – directing the hand to an object and guiding object manipulation in Land's terms (Johansson, Westling, Bäckström, & Flanagan, 2001). Such studies consistently show that actors predictively shift their gaze to upcoming contact events to guide the hand towards the target and typically maintain gaze at this location until the hand departs thus allowing contact events to be monitored (Flanagan, Bowman, & Johansson, 2006). Indeed, actors experience “gaze anchoring”, i.e. they find it difficult to shift their gaze away from a hand landing site before the arrival of the hand (Neggers & Bekkering, 2000; see also Ma-Wyatt, Stritzke, & Trommershäuser, 2010). The value that the oculomotor system places on fixating targets ahead of time is also reflected in low-level saccade metrics: saccades directed to a target in the service of eye-hand
coordination are slightly faster than predicted by the main sequence\(^2\) (Snyder, Calton, Dickinson, & Lawrence, 2002) in line with models of action control which predict increased movement vigor for actions of higher intrinsic value (Xu-Wilson, Zee, & Shadmehr, 2009).

1.2.3 Observer-Actor eye movement coupling in action observation

Crucially, chaining gaze behavior to the requirements of (hand) actions has turned out to be a key principle also for action observation. This was first demonstrated in a block stacking task, where participants showed the familiar predictive gaze patterns when engaging in the task themselves, but also exhibited qualitatively similar gaze behavior when observing an actor perform the same movement sequence (transporting three blocks one by one to the left and stacking them on top of each other, then replacing them back to the right – see figure 1.4b). Observers – like actors – shifted their gaze to the upcoming contact site (either the block to be grasped or the landing site where the block in hand would be placed) in time with movement onset, that is, in anticipation of the movement. Thus, the observer's gaze behavior is closely coupled to that of the actor in space and time, with the actor shifting their gaze on average slightly earlier than observers to the upcoming target location by around 100ms (Flanagan & Johansson, 2003 see figure 1.4).

---

2 Saccadic eye movements are highly stereotyped in that peak eye velocity more or less linearly depends on saccade amplitude for a given participant.
This finding was interpreted by the authors in terms of direct matching of observed actions with the observer's own motor programs otherwise used in eye-hand coordination when engaging in the task themselves. This interpretation is closely analogous to accounts of the mirror neuron system (Rizzolatti & Craighero, 2004; Rizzolatti & Sinigaglia, 2008). However, the exact relationship between the mirror

Figure 4 Gaze–hand coordination in action and action observation. 

**Figure 4** Gaze–hand coordination in action and action observation. **a, b,** Gaze positions at the end of periods between saccades (blue circles scaled to fixation duration), median hand path with and without a block in hand (unbroken and broken black lines, respectively), and approximate block positions before (right) and after (left) stacking. **c, d,** Median horizontal (x) positions of gaze (blue) and the index finger (black) as a function of time. Red traces represent the x position of gaze for all periods between saccades. **e, f,** Median vertical (y) position of the index finger. Broken and unbroken vertical lines indicate the times at which the index finger exited grasp and landing sites, respectively. The red arrow in **f** highlights that observers respond to the first transport action in the sequence with a short delay while subsequent gaze shifts occur in time with movement onset (second and thirst broken line). Figure and caption modified from Flanagan & Johansson (2003).

This finding was interpreted by the authors in terms of direct matching of observed actions with the observer's own motor programs otherwise used in eye-hand coordination when engaging in the task themselves. This interpretation is closely analogous to accounts of the mirror neuron system (Rizzolatti & Craighero, 2004; Rizzolatti & Sinigaglia, 2008). However, the exact relationship between the mirror
neuron system and the temporal coupling of observer and actor gaze described in the block stacking task has remained puzzling, despite their shared appeal to the direct matching hypothesis. In particular, F5 mirror neuron activity has an automatic character in that cells typically fire when observing a given hand action irrespective of attentional state, i.e. whether monkeys direct their gaze at the experimenter's action at all. This puzzle has only recently begun to be clarified (see Maranesi et al., 2013), see the detailed discussion in chapter 5, section 5.3.

1.3 Action Plans in Action Observation Revisited

1.3.1 Situating block stacking

The original demonstration of predictive gaze in action observation (Flanagan & Johansson, 2003) has been highly influential and is frequently cited by studies which make use of predictive gaze in observers. However, no single design is representative of the variety of contexts in which action observation occurs, and newer studies more often than not actually use designs and measures of “predictive gaze” which are quite different from the original 2003 paper which they nonetheless cite in support of their methodology while glossing over these conceptual and methodological differences. This will be discussed in some detail in chapter 5, section 5.2.

To get a deeper appreciation of the aspects of action observations captured in the Flanagan & Johansson (2003) study, and to what extent they may or may not generalize to other activities, it is essential to compare and contrast observer gaze behavior in a number of different activities. Unfortunately, we are currently lacking such studies, as naturalistic investigations are restricted to individualistic activities and laboratory studies on the whole either employ variations on the theme of block object manipulation, mostly using video clips, or still more artificial virtual designs. I will here first take a brief look into a controlled demonstration of action observation in a semi-naturalistic setting which we conducted as part of an exploratory study. To get insight into the relationship between actor and observer gaze behavior I constructed a parallel 3D eye tracking system in the laboratory for 2 participants.
wearing head-mounted eye trackers (ISCAN, Inc., Burlington, MA) and electromagnetic position sensors on their heads and fingers (Polhemus Liberty, Burlington, Vermont). From those data, a 3D model of their Lines of Sight (LoS) is created using a streamlined version of the 3D eye-tracking algorithms published by Ronsse and colleagues (2007). This 3D background model is then superimposed on a synchronized video stream recorded by an external camera mounted on a tripod using Augmented Reality overlays (Camera Calibration Toolbox, see Bouguet, 2008). This setup provides automatic detection of eye hand coordination and social gaze coordination between actor and observer and clear visualization (as opposed to relying on 2 streams of shaky head cameras).

Kelly Woudsma and I looked at how participants would naturally interact with and observe each other's actions. We designed a LEGO model train building task which we made incrementally more stringent e.g. by standardizing positions of all tiles to bring it closer to the original block stacking design. Participants first observed and then constructed the model train themselves and did so twice to investigate effects of familiarity and practice, with the whole procedure taking about 20 minutes. A representative section of 15 second duration is shown in figure 1.5.
Figure 5 The demonstrator (sitting on the right) is building a modular LEGO model train, with colored tiles prepared in arrays on the table. The observer (left) has already observed and performed the construction of the model train once. Here the demonstrator is reaching with her right hand for a yellow tile from the first array, then retrieves it and stacks it together to build the first cart which is done bimanually. Six tiles are retrieved so this sequence is repeated 6 times. Colored traces show hand and gaze measures for the demonstrator and the observer. From top to bottom: The demonstrator's right hand entering the area around the array of yellow bricks (6 yellow traces on top, shown against gray background); the right hand returning back to her home workspace (magenta); as well as her left hand (which stayed in the home zone throughout the construction). The next 2 traces constitute measures of eye-hand coordination showing the demonstrator's gaze when within 3cm of her right (blue) and left (red) index fingers. The equivalent measures are shown for the observer (second blue and red traces). The lowermost black trace denotes shared gaze events defined as the demonstrator and observer's lines of sight being less than 3cm apart. The 4 black arrows highlight the observer lagging significantly behind the actor when monitoring her right hand retrieving a new tile. This is shown in more detail in the 4 still frames taken from the time period indicated by the black rectangle: both are monitoring the stacking together of the first 2 tiles (thick lines connecting eye and table intersect in magenta), whereas the demonstrator first shifts her gaze (2nd still image) and
As in the original study (Flanagan & Johansson, 2003), almost all of the observer's fixations are directed at contact events and the observer's gaze follows the actor's out and back hand movements to the tile arrays as well as the bimanual assembly of tiles into train carts. Apart from this there are only rare additional saccades to the finished train carts presumably to help with memorizing the required arrangement of the LEGO pieces. An impressive number of shared gaze events occur, some of extended durations, suggesting that both actor and observer are similarly engaged in the task. However, the observer's gaze is regularly initiated in reaction to, rather than in time with, the onset of the actor's reaching out to retrieve the next tile, thus lagging behind the gaze of the actor - even though the tile pickup sites have been standardized and the observer at this stage is highly familiar with the action sequence and their roles have been reversed twice. What accounts for these differences in observer gaze behavior between the block stacking and the LEGO building task?

1.4 An organizational framework of predictive gaze in action observation

1.4.1 Modes of predictive gaze

Clarifying the concept of predictive gaze by introducing distinct modes of predictive gaze behavior in action observation constitutes the first of two conceptual contributions introduced here. The separate modes are outlined in figure 1.6 and described below (for detailed substantiation and a demonstration of how latency criteria can be derived the reader is referred to chapter 4). These distinctions in terms of the nature of the predictive processes on the part of observers are key and may even seem obvious in hindsight but their recognition in the literature is uneven (this is discussed in detail in section 5.2.2).
1.4.1.1 Anticipation

Anticipatory predictive gaze shifts to an estimated upcoming target location are defined here as being initiated proactively based on the observer's prior estimate – as opposed to generated in response to a perceived event. Anticipatory predictive gaze shifts may be initiated in advance, in time with, or slightly after\(^3\) hand movement onset. As gaze shifts are generally much faster than hand movements, anticipatory gaze shifts typically arrive at a target earlier than the hand by a considerable fraction of hand movement time (figure 1.6, red trace). This is the mode of predictive gaze that is typically found for actors, and was described in the original report for observers as well (Flanagan & Johansson, 2003). Being based on an estimate, the observer in principle has no guarantee that the actor is actually about to initiate a movement.

3 Gaze shifts are categorized as anticipatory as long as the observer initiates her movement before sensory information about the the actor's movement onset is available and can effectively be used to guide gaze shifts. Beyond this point, gaze shifts could potentially have been generated in response to the action and thus should be categorized as based on extrapolation instead.
to the estimated location at this time (i.e. the action may be delayed due to the actor having trouble completing the previous action).

1.4.1.2 Extrapolation

After movement onset, online visual cues become available indicating 1) that an action has been initiated and 2) the general direction of the trajectory of the action which can be used to disambiguate actions to different potential targets. Gaze shifts which are initiated in response to movement onset but overtake the hand en route and thus arrive at the target ahead of time are categorized as predictive gaze based on extrapolation, as illustrated in figure 1.6 (blue trace). The maximum latency that leaves just enough time for gaze to overtake the hand/cursor and arrive at the target ahead of time will depend on the relative movement duration of the action compared to the observer's gaze shift.

1.4.1.3 Tracking

Technically, tracking the hand rather than moving ahead to a likely target also has to rely on prediction since a purely reactive process would lag behind in time due to sensory and motor delays (e.g. Krauzlis, Liston, & Carello, 2004). Episodes of tracking may occur especially for slow movements and observers may initially track an action trajectory before shifting their gaze ahead to the target based on extrapolation. Thus (partial) tracking may result in the observer's gaze arriving at the target location in advance of the actor's hand and care needs to be taken to identify episodes of tracking in a dataset.

1.4.2 A framework of factors determining modes of predictive gaze

As a second conceptual contribution I introduce an organizational framework of factors which influence the mode of observer gaze behavior in action observation. The framework is organized around 4 questions, all of which need to be considered at minimum to make sense of gaze behavior in action observation in a given activity and what degree of actor-observer gaze coupling can be expected (see table 1.1). Some of these factors are associated with characteristics of the actor - the ease with which their movements can be read, their movement kinematics, and their goals - but also crucially depend on
characteristics of the observer - their knowledge and skills, their perspective, as well as the observers' own goals and interests. Another key influence on action predictability is the spatial configuration of the scene, especially locations of potential target objects.

From this perspective, part of what made the original block stacking design (Flanagan & Johansson, 2003) so inspired was keeping all these aspects as similar as possible between actor and observer. First, the movements of block stacking are executed in a stereotyped, rhythmic manner, making the onset times of upcoming actions in the sequence transparent to the observer. Second, the scene and movement metrics were kept identical between actor and observer, as participants adopted the same position in the fixed eye

<table>
<thead>
<tr>
<th>Timing: When?</th>
<th>Actor</th>
<th>Observer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Readability</td>
<td>Familiarity:</td>
</tr>
<tr>
<td></td>
<td>Rhythmicity</td>
<td>Knowledge and skills pertaining</td>
</tr>
<tr>
<td>Scene Configuration: Where? &amp; Movement Metrics: How?</td>
<td>Hand and eye movement opportunities afforded by the scene from the actor's perspective</td>
<td>Eye movement opportunities afforded by the scene from the observer's perspective</td>
</tr>
<tr>
<td>Task Engagement: Why?</td>
<td>Actor's goals:</td>
<td>Observer's interests and goals:</td>
</tr>
<tr>
<td></td>
<td>Guiding/Directing</td>
<td>Monitoring</td>
</tr>
<tr>
<td></td>
<td>Monitoring</td>
<td>Updating</td>
</tr>
<tr>
<td></td>
<td>Updating</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Framework of factors determining modes of predictive gaze behavior in action observation of the action and of the scene, the observer's perspective, their familiarity with the activity, and the motivations of both parties.
tracker for both actor and observer trials, and while the hand approached the blocks from different angles, the endpoint kinematics in actor and observer trials were essentially equivalent (Flanagan & Johansson, 2003, see figure 1.4). Third, observers were highly familiar with the block stacking sequence, having themselves performed it at length. Finally, I suggest that the action goals for gaze shifts in eye-hand coordination on the part of the actor – directing and guiding movements to contact events while monitoring task progression – are partially shared by observers. This perhaps counter-intuitive notion is developed in more detail below.

In contrast, the more complicated LEGO train construction task – and perhaps (semi-)naturalistic settings more generally – is much less transparent to observers as the duration necessary to successfully stack LEGO tiles is much more variable even for trained builders (compare timing of hand movements between figure 1.4c,e and figure 1.5, top 2 traces). This makes it difficult for the observer to predict with certainty when the actor will reach for the next tile and to initiate her gaze shift to the upcoming target site in time with movement onset, even though the upcoming target location is known in advance. Moreover, the actor routinely performs more than one action at a given time, i.e. reaching for the next tile with the right hand while finishing clicking the current tile into place with the left hand, thus introducing gaze competition between contact events occurring concurrently at different locations in the scene. This presence of multiple “legitimate” gaze targets further invites variability between actor and observer gaze.

The proposal made here is that we can expect the close coupling in space and time between the actor and observer's gaze (i.e. anticipatory predictive gaze) to begin to break down as soon as an (experimental) situation deviates from providing the equivalence between actor and observer inherent in the original block stacking design along any of the aspects outlined in the framework. This will occur when the observer's insight into the activity varies from the actor's, multiple potential action targets are presented, observers pursue interests different from actors, etc. As each of the factors outlined in the framework may shape the mode of gaze behavior adopted by the observer – as will be demonstrated in the following section – gaze behavior does not necessarily directly reflect prior knowledge and motor skills of observers. Instead, all the factors outlined in the framework need to be considered for interpreting gaze in
action observation. In addition, as observer gaze becomes systematically de-coupled from the actor's eye-hand coordination, we should also expect observer gaze behavior to become much more variable (both between participants as well as within different trials of the same participant) as the uniformity (the “clock-work” character) of gaze behavior documented in action observation for block stacking is certainly the exception in gaze allocation more generally (see section 1.2).

1.4.3 Prior Knowledge of Actions, Scene and Movement Metrics, and Observer Goals as Determinants of Gaze Behavior in Action Observation

For observers to be capable to regularly shift their gaze to an upcoming contact event in time with the actor's movement onset (anticipatory predictive gaze), they, like actors, need to have a detailed understanding of the activity in terms of action timing and target locations (compare Land & Furneaux, 1997). The predictability of actions for observers has been systematically varied to explore the consequences on gaze behavior in action observation in follow-up studies to the original report by Flanagan & Johansson (2003) which I review below to flesh out the organizational framework introduced above in more detail.

Knowing When: Due to its regular and rhythmic movement sequence, observers in the block stacking study could achieve action timing estimates comparable to those of the actor. However, observers' gaze exited the first grasp site with a slight delay, interpreted as observers having to get attuned to the actor's rhythm (see figure 1.4d, red arrow). This underscores that for observers to commit to initiating an eye-movement in anticipation of an action they need to be able to estimate with sufficient certainty when an upcoming movement is executed, even when the type and spatial target of the upcoming action is known in advance. Otherwise gaze shifts will likely be initiated in response to perceiving the onset of the action (predictive gaze based on extrapolation of an ongoing action).

Knowing Where: The influence of prior knowledge of the upcoming grasp target location has been investigated in detail in an object lifting task by Rotman and colleagues (2006). In their first experiment, illustrated in figure 1.7a, the actor initially lifted the middle (start) object (indicated by the arrow), and
then proceeded to grasp and lift one of the two candidate objects located to the left or to the right (first lift, unpredictable location), before lifting the start object again followed by lifting the remaining target object (second lift, predictable location).

Figure 7 Design in the block lifting experiments used by Rotman et al. (2006) shown from the observer's point of view. Black arrows indicate the start block. a Target blocks located to the left and right of the start block (located further away). b Near (middle) and far (left) target blocks are located along a line.

Observers in almost all trials looked exclusively at the objects that were lifted. When observing predictable movements, observers initiated their gaze shifts from the start to the correct target block in anticipation of the movement with mean gaze latencies of 77ms after movement onset. While slightly longer than the latencies reported for observers in the original block stacking experiment (28ms after movement onset), such gaze shifts still clearly reflect anticipatory initiation, as these intervals do not leave enough time for the oculomotor system to effectively integrate visual information for saccade planning (Desmurget & Grafton, 2003). Conversely, when observing unpredictable actions, observers had to rely on visual inspection and extrapolation of the actors movement trajectory to correctly estimate the target in each trial. Observers accomplished this quite quickly, shifting their gaze away from the start target on average by 156ms after hand movement onset. In both cases, observers' gaze typically arrived at the target objects ahead of the actor's hand, i.e. gaze was “overtaking” the actor's hand en route to the target due to the faster eye movements compared to hand movements. Thus while observers incurred a delay in unpredictable trials (where they had to rely on extrapolation) compared to predictable trials
(anticipation), they were still able to monitor hand-object contact and object placement events with central vision.

In a second experiment carried out by Rotman et al. (2006), objects were arranged along a line (see figure 1.7b). As before, the actor began a trial by lifting the start block (see arrow), then proceeded to lift either the near or far block (first lift, unpredictable location), then lifted the start block again followed by the remaining target block (second lift, predictable location). For this target configuration observers almost always shifted their gaze away from the start block in anticipation of the hand movement (on average 54ms after movement onset). This was the case even when observing unpredictable actions, where participants in most trials shifted their gaze to the near target also in anticipation of the action, and then generated a second, corrective saccade when the actor reached for the far block first. In fact, even when actions directed at the far target were predictable, observers shifted their gaze directly from the start target to the far target in only about half of the trials, whereas in the other half they still briefly shifted their gaze to the near target first before moving ahead to the far target as the action progressed. Thus the two scene configurations tested (see figure 1.7) resulted in quite different sets of observer gaze behaviors. In particular, the results in unpredictable trials for the second configuration indicate that anticipatory gaze shifts to a specific target object do not necessarily reflect prior knowledge on the part of the observer, but may instead reflect the “strategic value” of the object's location within the scene, which may be closer to the current gaze target to begin with, located near other potential targets, and offering a convenient vantage point for monitoring the actor's approach. The relative influence of prior knowledge of actor target selection and scene configuration on observer gaze behavior is investigated in more detail in chapter 4.

**Observer Goals:** Finally the observer's goals may constitute a critical determinant of gaze behavior. The close correspondence between observers' and actors' gaze behavior in block stacking (Flanagan & Johansson, 2003) itself suggests that the actor's goals of visually supporting the guiding, directing, and monitoring of actions (Land, 2006) are also at least partially shared by observers. This proposal may be somewhat counter-intuitive: actors have every incentive to be vigilant in monitoring their actions and not
“take their eye off the ball”, whereas observers surely are not actually at risk of fumbling the ball or spilling their tea⁴. Still, there are other aspects of action planning and control which may be naturally of interest to observers as well. These include updating their understanding of developments in the world by monitoring the actor's task progression and checking for unexpected occurrences, thus potentially learning about object properties, about the sequential structure of the activity, and about dispositions and propensities of actors. This is in line with the general outlook in motor control which regards constant adaptation and learning as an intrinsic part of action planning and control (Shadmehr & Wise, 2005; Wolpert, Diedrichsen, & Flanagan, 2011).

In particular, observers may learn about, keep track of, and adapt their estimates of object properties via action observation. In this observers, like actors, may benefit from monitoring informative events such as object contact, lift-off, and placement, especially with regard to object properties such as weight which they cannot deduce with certainty from visual inspection of the object at rest (Flanagan et al., 2006). To clarify the relationship between the different functions proposed for gaze behavior in action observation, Flanagan et al. (2013) explicitly instructed observers either to predict which of two candidate objects the actor would lift (target prediction task) or judge the weight of the object that the actor selected and lifted (weight judgment task). Gaze behavior in these two tasks was then compared to uninstructed gaze behavior (self-guided action-observation) similar to Yarbus' (1967) methodology (see figure 1.3). The design used a similar target configuration to the one illustrated in figure 1.7b, with target locations being always unpredictable. In the weight judgment task, the target blocks were attached to weights, located below the table surface, via strings and the weight of the object lifted was occasionally increased. Observers had to press one of two buttons (indicating light weight or increased weight) as quickly as possible. In this task, observers' gaze behavior was qualitatively identical to the free-viewing condition,

⁴ Some authors disavow ascribing functionality to proactive observer gaze behavior altogether: “There is no direct rationale for the proactive gaze shifts performed during action observation” (Rosander & von Hofsten, 2011).
except that observers initiated their anticipatory gaze shifts (directed towards the near block) even earlier presumably indicating high motivation due to the explicit task. In the target prediction task, observers were instructed to press one of two buttons indicating their target estimate (near or far block) as quickly as possible. In this condition, most observers qualitatively changed their gaze behavior as compared to free viewing: rather than shifting their gaze rapidly to the object grasp sites, 5 out of 10 participants instead fixated exclusively on a central location in most trials, and 3 more participants switched to a tracking strategy. These results suggest that predicting the upcoming location of an action by itself is unlikely to be the primary goal of observers since when target prediction is reinforced through instructions they readily shift to gaze strategies rarely seen in free viewing. Instead, we propose that observers predict upcoming target locations as a means to direct their gaze to contact events ahead of time in order to monitor activity progression and extract information about objects, including object weight, using central vision.

1.5 From Action Observation to Social Learning in Action

The studies of gaze behavior in action observation described in the previous section have led us to conclude that some of the concerns of the actor – monitoring contact events and updating estimates about objects in the service of guiding object-oriented movements – are also likely shared by the observer. Thus action observation becomes intrinsically linked with social (motor) learning. Elucidating this relationship is the main drive behind the studies reported in this thesis (chapters 2-4, see below).

Social learning by observing has been studied by a number of fields. Within psychology, comparative and developmental psychology have focused on learning of arbitrary stimulus-response relationships (e.g. pressing a lever to release food, finding food at a specific location), or to open a box containing food using a particular effector, i.e. paws or mouth (Heyes & Galef, 1996; Huber, 2012). In most if not all of these cases, learning is conceptualized as having an all-or-none character, although animals may learn only particular aspects or exhibit behavior patterns of their own in addition to what they have learned from observation.
In contrast, learning in tasks studied in motor control has a more incremental nature and may be conceptualized as adaptation learning. For example, an influential design consists of planar reaching tasks where participants hold a manipulandum connected to a robot which exerts forces on their hand as a function of position or speed, etc. (“force field” adaptation learning). Fully (re-)learning to reach quickly and accurately in the presence of such perturbations may take on the order of hundreds of trials (Shadmehr & Mussa-Ivaldi, 1994). Whereas learning to smoothly lift a novel object is accomplished much more quickly, after only a few or even a single lift (Johansson & Westling, 1988), lifting performance (scaling of fingertip forces) is still more or less complete and needs to be assessed quantitatively.

Force field adaptation learning by observing has been described in extensive detail: when participants first get to passively observed someone else learning to reach (i.e. seeing pre-recorded cursor motions from other participants) in a novel mechanical environment (in the presence of a clockwise or counterclockwise force field), they then show extensive savings when learning to reach in the same conditions themselves afterwards (Mattar & Gribble, 2005; Malfait et al., 2010). This line of research can be regarded as setting the standard for investigations of motor learning by observing. However, reach adaptations in novel mechanical environments likely tap into corrective processes which normally allow us to constantly keep track of and adapt our motor system to changes in the periphery (i.e. muscles becoming tired, injury), making it a model system to study motor control and motor learning in the laboratory. In contrast, it is far less clear what social aspects of learning by observing are captured when this design is extended to learning by observing. Effectively, social motor learning becomes operationalized as a passive visual experimental condition. While this certainly captures some aspect of social learning by observing – which had before been absent in psychological studies, see above – this represents a very narrow view on action observation and social engagement and it remains to be seen whether and how people in real life contexts use this ability to learn novel mechanical environments drawing on someone else's experience. In contrast, learning by observing about object properties such as weight – while also a narrow but well-controlled aspect of social learning more generally – at least offers
much more evident and plausible advantages in real life settings, as we regularly see other people manipulate object whose weight cannot always reliably inferred from static visual cues alone.

1.5.1 Chapter 2

In this study we describe object lifting as a model system for social learning about object weight. As the weight of objects at rest cannot always accurately be inferred from visual cues, there is an incentive for observers to monitor actors manipulating objects, in particular contact events, to learn about and keep track of the weights of objects around them. Here we test the proposal that observers, like actors, use those predictions – in particular prediction errors – to update their estimates about object weight to be more effective in lifting these objects themselves. While there have been a number of studies showing that observers can report object weight (Runeson and Frykholm 1981; Bingham 1987; Shim and Carlton 1997) – and even a number of studies describing neural correlates of observing an actor lift an object (see Alaerts et al. 2010) – this is the first study to explicitly address how social learning actually translates into lifting performance on the part of an observer. Participants took turns lifting an object with a demonstrator while the object's weight was changed in blocks between light and heavy. The weight always changed for the demonstrator first, allowing us to quantify precisely how learning by observing translated into updating of participants' fingertip forces on the object and thus their lifting performance.

1.5.2 Chapter 3

As indicated in the exploratory studies shown in figures 1.1 and 1.5, multiple actions may unfold concurrently in complex activities, creating competition for the observer's gaze. Here we create a model system where a demonstrator simultaneously lifts two objects to test the proposal that observers selectively allocate their attention to objects to extract information relevant to guiding their own actions. In chapter 2, we show that observers watching an actor lift a single object intuitively pick up on weight changes through observation to guide their own subsequent lifting actions in the absence of explicit
instruction. We propose that this social (motor) learning is in part what drives the observer's gaze to object contact events in action observation.

In a given trial of the task used in the study described in chapter 3 the demonstrator first lifts a central start object with both hands and then reaches outwards with each hand to simultaneously lift two target objects located on either side of the central start object. Participants are then cued to lift one of the target objects. The weight of one of the target objects was occasionally set to heavy at the start of the trial, so that participants could again benefit directly from observing the demonstrator's action. To test the hypothesis that participants preferentially shift their gaze to objects whose properties are relevant to guiding their own actions, we made one target object more likely for them to interact with subsequently by repeatedly cueing each target objects across a block of trials in experiment 1, and made one target object more challenging to lift by changing the weight only for this object in experiment 2.

By measuring how participants allocate their gaze to the two target objects as they are being lifted by the demonstrator, as well by measuring their fingertip forces when the participant lifted the cued object themselves afterwards, this design allows us to address causes and consequences of gaze behavior in action observation. We expected gaze to be directed preferentially to the object-to-be-lifted (experiment 1), and the the object with variable weight (experiment 2).

1.5.3 Chapter 4

How does knowing the circumstances of an action translate into predictive gaze behavior in action observation? Here we recorded eye movements of participants watching an actor engage in choice behavior in a virtual target hitting task. The actor first scored a start target by moving and holding a cursor with a joystick, and then hit one of two choice targets whose positions were systematically varied. In the 5:5 condition, both choice target were worth 5 points to the actor and in the 9:1 condition, one colored target was worth 9 points and the other worth 1. The observer was explicitly told these target values in these two condition. In the baseline condition, the actor was told which choice target to move to and this target was randomly selected, making it impossible for the observer to correctly predict that choice target.
By describing in detail how the observer's knowledge – or lack thereof – about the actor's target preference combined with the target configuration translates into observer gaze behavior, and characterizing gaze strategies and evaluating their visual consequences, this study lays the groundwork for investigating of social learning of object value through action observation from a motor control perspective.

1.6 References


29


Chapter 2

Adaptation of lift forces in object manipulation through action observation

2.1 Abstract

The ability to predict accurately the weights of objects is essential for skilled and dexterous manipulation. A potentially important source of information about object weight is through the observation of other people lifting objects. Here we tested the hypothesis that when watching an actor lift an object, people naturally learn the object's weight and use this information to scale forces when they subsequently lift the object themselves. Participants repeatedly lifted an object in turn with an actor. Object weight unpredictably changed between 2 and 7 N every 5th to 9th of the actor’s lifts and the weight lifted by the participant always matched that previously lifted by the actor. Even though the participants were uninformed about the structure of the experiment, they appropriately adapted their lifting force in the first trial after a weight change. Thus, participants updated their internal representation about the object's weight, for use in action, when watching a single lift performed by the actor. This ability presumably involves the comparison of predicted and actual sensory information related to actor’s actions, a comparison process that is also fundamental in action.

2.2 Introduction

Skilled object manipulation, including tool-use, requires learning object dynamics, which specify the relation between object motion and applied force. Such learning is considered to involve two components: learning the structure of the dynamics, captured by the form of the equations of motion, and learning the parameters for a given structure (Braun et al. 2009; Braun et al. 2010). Many studies of motor learning have examined point-to-point movements of a grasped object with novel and unusual equations of motion, often imposed by a robot (Lackner and DiZio 2005; Shadmehr et al. 2010; Wolpert and Flanagan 2010). In such cases, in which the actor must discover the structure of the dynamics, learning typically
requires tens to hundreds of movements. However, most of our skill learning involves learning parameters of familiar dynamics, such as the weight of a book or the stiffness of an elastic band, and typically occurs over one or a few trials (Johansson and Westling 1988; Flanagan and Wing 1997; Ingram et al. 2010).

Although the adage practice makes perfect certainly applies to skill learning, the observation of others’ actions constitutes an important source of information for such learning. It is well established that high-level task information, such as the sequence of required movements, can be learned through action observation (Heyes and Foster 2002; Torriero et al. 2007). However, recent findings indicate that action observation can implicitly facilitate learning of the structure of dynamics of novel loads (Mattar and Gribble 2005; Brown et al. 2010). Specifically, watching a video of an actor learning to move an unusual hand-held load that initially perturbs hand motion leads to some improvement of skill acquisition when the observer subsequently performs the same skill-learning task.

In the current study, we examined whether people also acquire knowledge about the parameters of familiar loads through action observation. Using an object-lifting task we tested the hypothesis that people naturally update knowledge related to object weight, used to adapt force output when subsequently lifting the object, based on watching an actor lifting the object. Previous studies have examined the effect of action observation on judgements of heaviness (e.g., Hamilton et al. 2007) and how heavy something looks (Runeson and Frykholm 1981; Bingham 1987; Shim and Carlton 1997). To our knowledge, only one study has examined the effect that action observation has on lifting behaviour (Meulenbroek et al. 2007). In the latter study, in each trial one participant lifted and then placed an object in a shared workspace and a second participant then lifted and retrieved it. The authors examined trials in which object weight unexpectedly changed and compared the change in lift height, relative to the previous trial, in placers and retrievers. As expected, in trials in which object weight decreased, lift height increased in placers. A slightly smaller increase in lift height was seen in retrievers, indicating a weak effect of action observation. Similarly, when object weight increased, a decrease in lift height was observed in both groups but this decrease was slightly smaller in retrievers than placers. However, the latter result is less
clear-cut because placers and retrievers exhibited different lift heights in the trial prior to the weight increase.

Here we assessed, in addition to lift height, the forces applied to the object prior to lift-off, which, in trials in which weight increases, provide an earlier and clearer estimate of expected weight that is not affected by corrective mechanisms. Participants repeatedly lifted an object in alternation with an actor. The weight changed, unpredictably, every 5th to 9th of the actor’s lifts so that participants (and the actor) could not reliably predict when the change in weight would occur. However, because the weight lifted by the participant always matched that of the actor, the participant could potentially gain information about the weight of the object in their forthcoming lift by observing the actor’s current lift. We found that even though participants were naïve about the structure of the experiment, they effectively adapted the lifting force when lifting the new weight. Thus, participants exhibited rapid, single trial parametric learning through action observation.

2.3 Methods

2.3.1 Participants and general procedure

Nine participants, including 6 women and 3 men, were recruited from the population of undergraduate and graduate students at Queen’s University. Participants provided written informed consent and received monetary compensation for their time. The ethics committee of Queen’s University approved the study. For the analysis of lift height (see below), one participant was excluded because of missing object position information.

Seated participants repeatedly lifted an object located on a tabletop in front of them either by themselves or in turn with an actor seated on the other side of the table (figure 2.1A). Each lift was initiated by a tone that instructed the participant or actor to grasp the object by the handle using a precision grip and lift it. A second tone delivered 750 ms after lift-off instructed the lifter to replace the object on the table in the same location. Participants, and the actor, were encouraged to lift the object
smoothly to a height of approximately 2 cm and to keep the duration and height of the lift consistent throughout the experiment. Participants and actor used the whole arm when lifting (i.e., they did not just lift via wrist movement) and rested their hand and forearm on the table between lifts. The time between successive lifts was approximately 5 seconds.

Figure 2.1 Experimental set-up and data analysis. A: Top view schematic showing the positions of the actor and participant in conditions in which they lifted the object in alternation. B: Object with handle instrumented with force sensors. C: Diagram of the linear motor system used to programmatically specify object weight. D: Load force (LF) and LF rate functions from two lifts of the 7 N weight. In one lift (grey curves) the initial increase in load force undershot object weight and, in the other lift (black curves), the initial increase in LF was close to the mark. Note that the initial peak in LF rate scaled with the initial increase in LF, which depends on expected object weight.

One of the authors (A. Reichelt) served as the actor throughout the experiment. Importantly, because we randomly changed the sequence of object weights for each experimental session (i.e., participant), the actor could not predict when changes in object weight would occur. Moreover, as will be described in the Results, analysis of the actor's lift forces confirmed that the actor did not predict weight changes. The
actor was instructed to maintain the same expression throughout each session and not to express surprise, either verbally or facially, during lifts when the weight changed.

2.3.2 Apparatus

The object consisted of a 5 cm³ hollow cube made from the opaque black polyoxymethylene plastic Delrin® (figure 2.1B). A handle mounted on the top of the cube included two force-torque sensors (Nano 17 F/T, ATI Industrial Automation, Garner, North Carolina) that measured the forces applied by the tips of the thumb and index finger. A flat circular disk, covered by medium-grain sandpaper, capped each sensor. The two disks, and hence the surfaces contacted by the thumb and index finger, were oriented in parallel vertical planes, separated by a distance of 4 cm. A miniature electromagnetic position sensor (Polhemus Liberty, Burlington, Vermont), attached to the side of the object, measured the height of the lift.

The weight of the object was set to either 2 or 7 N (0.196 or 0.687 kg) on a given lift and could be changed between lifts without the knowledge of the participant or the actor. The change in weight was implemented by a linear motor programmed to position a trolley along a rotating rod attached, via a string, to the center of the object (figure 2.1B). The string passed through pulleys and through a small hole in the tabletop, to a hook located in the center of the cube, the bottom of which was open. The trolley moved between every trial to guard against the lifter using the sound of the linear motor system as a cue signaling a weight change.

2.3.3 Experimental conditions

2.3.3.1 Solo condition

In this condition, participants repeatedly lifted the object themselves. Without changes of the visual appearance of the object, its weight switched between 2 or 7 N across blocks of lifts, starting with the 2 N weight in the first block. The number of lifts per block was randomly varied between 5 and 9. Thus, the participants could not reliably predict when the weight would change based on the number of lifts. The
participant completed 12 blocks of lifts. This provided 6 transitions from the 2 N weight to the 7 N weight. The solo condition provided a baseline with which to compare with the other conditions.

2.3.3.2 Coupled condition

In this condition, the actor and participant performed alternating lifts of the object with the actor going first. For the actor, the weight switched between 2 and 7 N across 12 blocks of 5 to 9 lifts, starting with the 2 N weight and the participant experienced the same sequence of weights. Thus, neither the participants nor the actor could reliably predict when the weight would change based on the number of lifts.

2.3.3.3 Informed condition

This condition was the same as the solo condition, except that before each weight change, the experimenter verbally informed the participant about the change and indicated whether the new weight was light or heavy. Specifically, the participant was told either “the weight has changed and is now light” or “the weight has changed and is now heavy”. The informed condition allowed us to compare scaling of load forces on transition lifts in the coupled condition with that occurring when participants were explicitly informed about weight changes.

Each participant first completed the solo condition (solo1), followed by the coupled and informed conditions in counterbalanced order. They then completed a second solo condition (solo2). Because the solo1 condition was performed first, participants had experienced both weights and had learned that object weight could change when performing the coupled and informed conditions. The solo2 condition allowed us to evaluate the consistency of the participant’s behavior during the experimental session by comparing the lifting performance before and after the coupled and informed conditions.

2.3.4 Data Analysis

Position and force signals were sampled at 1 kHz and smoothed using a fourth-order, zero-phase lag, low-pass Butterworth filter with a cutoff frequency of 14 Hz. Note that the position signal was updated at
240 Hz and therefore this signal was oversampled. The vertical load force (LF) applied to lift the object was computed as the sum of the vertical forces applied to the opposing contact surfaces of the handle and the grip force was computed as the average of the normal forces applied to the surfaces. The rate of change of LF with respect to time, or LF rate, was computed using a first order central difference equation. Our analysis focused on initial lifts of the 2 N and 7 N weights that followed blocks of 7 N and 2 N lifts, respectively. However, for reasons outlined below, we used different measures to assess lift performance for the 2 and 7 N weights.

2.3.4.1 Analysis of 7 N lifts

Under conditions in which object weight is accurately predicted, people tend to lift objects of varying weight in about the same amount of time. To accomplish this, they scale the LF rate, prior to lift off, to the expected weight of the object – increasing load force more rapidly for objects they expect to be heavy. In addition, people predict the LF required for lift-off and, after initially increasing LF rate, they reduce LF rate so that it approaches zero at the expected lift-off time. In general, the peak rate of change of LF during the initial increase in load force, which we will refer to as the initial peak LF rate, provides an index of predicted weight (Johansson and Westling 1988; Flanagan and Beltzner 2000; Flanagan et al. 2008; Baugh et al. 2012). However, if the object is far lighter than expected, there may be an abrupt cessation in LF increase at the time of lift-off. If lift-off occurs before the time at which the initial peak LF rate would have occurred had the object been as heavy as expected, the measured peak LF rate may not provide an accurate index of the expected weight (for further details see Johansson and Westling 1988). For this reason, we used the initial peak LF rate to examine the first 7 N lifts (following each block of 2 N lifts), but did not use this measure to examine the first 2 N lifts (following each block of 7 N lifts). In addition to determining the initial peak LF rate, we quantified the duration of the load phase – during which LF increases before lift-off – as the time from when LF exceeded 0.05 N until the time LF reached the weight of the object (see figure 2.1D).
For comparison with the first 7 N lifts, we also analyzed the second and last 7 N lifts of each block of lifts of the 7 N weight as well as the last 2 N lifts that preceded the first 7 N lifts. Based on previous work showing rapid updating of load forces across lifts when a prediction error occurs (Johansson and Westling 1988; Gordon et al. 1993), we expected smaller prediction errors on the second 7 N lifts, in comparison to the first 7 N lifts, and the most accurate predictions of the current weight in the last 7 N lifts.

For completeness, we also measured the initial peak rate of change of grip force (GF), which we will refer to as the initial peak GF rate. Because required GF depends on LF, and is generally modulated in synchrony with LF (Johansson and Westling 1988; Johansson and Flanagan 2009), we expect the analyses of GF and LF to reveal similar results. However, because GF also depends on factors other than LF, including the frictional conditions between the digits and contact surface and the GF safety margin selected by the lifter, measures based on GF do not provide a direct measure of expected weight.

2.3.4.2 Analysis of 2 N lifts

When lifting objects that are lighter than expected to a small height (e.g., 2 cm as in the current study), people tend to overshoot the target height (Johansson and Westling 1988). Therefore, we used the maximum lift height to examine the first 2 N lifts (following each block of 7 N lifts). For comparison with the first 2 N lifts, we also analyzed the second and last 2 N lifts of each block of lifts of the 2 N weight as well as the last 7 N lifts that preceded the first 2 N lifts. We expected that lift height would decrease on the second and last 2 N lifts, in comparison to the first 2 N lifts, as participants update their prediction of object weight.

To assess the effects of condition and lift type (i.e., first, second, and last lifts), we used repeated-measures ANOVA as well as planned within-subject comparisons. The Holm-Bonferroni test was used to correct for multiple comparisons. This test fully controls for family-wise error, but is more powerful than the stringent Bonferroni test. A p-value of 0.05 was considered statistically significant.
2.4 Results

2.4.1 Single trial results

Figure 2.2 shows load force (LF), LF rate, and lift height (i.e., vertical positions) records from single trials, performed by a single participant, from the solo1, coupled and informed conditions. Very similar force and position records were observed, across conditions, for the last lifts in the blocks of lifts with the 2 N and 7 N weights. This result is expected because, in all three conditions, participants could rely on sensorimotor memory of the previous 4-8 lifts with the same weight in order to scale load force to the current weight. In fact, similar force and position records were also observed for the second 2 N and 7 N lifts, consistent with previous work showing that rapid, single-trial updating of sensorimotor memory for weight (Johansson and Westling 1988; Johansson and Flanagan 2009).
Figure 2.2 Load force (LF), LF rate, and lift height records from a single participant. The top row shows examples of last lifts of a block of lifts of the 2 N weight and the first and second lifts of the subsequent block of lifts of the 7 N weight. The bottom row shows examples of last lifts of a block of lifts of the 7 N weight and the first and second lifts of the subsequent block of lifts of the 2 N weight. The records are colour coded by condition (see inset).
The key trials are the first 7 N and the first 2 N lifts that followed blocks of 2 N and 7 N lifts. For the first 7 N lifts, the force records differed considerably across conditions for this participant. As expected, in the solo1 condition, the object did not lift off after the initial increase in LF and an additional LF increase was required to achieve lift-off. However, in the coupled and informed conditions, the initial increase in LF was sharper and reached a greater force such that an addition LF increase was not required. Thus, this participant predicted the object weight quite well immediately after the weight transition in the coupled and informed conditions, but not in the solo1 condition. For the first 2 N lifts, the lift height records differed considerably across conditions. As expected, in the solo1 condition, a strong overshoot in lift height was observed, indicating that the participant was fooled by the decrease in weight. However, in the coupled and informed conditions, the maximum lift height was similar to that seen in the second and last lifts of both the 2 N and 7 N weights, indicating that participants predicted the weight quite well.

The left panels of Figure 2.3 show, for each condition, cumulative distributions of initial peak LF rates across the six first 7 N lifts performed by each participant in each condition. The right panels of Figure 2.3 show corresponding distributions of maximum lift height across the six first 2 N lifts. On average (see grey vertical lines), peak LF rates in 7 N lifts were greater, and maximum lift heights in 2 N lifts smaller, in the coupled and informed conditions compared to the two solo conditions. Of particular interest is whether, in the coupled condition, the apparent benefits of observing the actor are sporadic (i.e., seen in some first 2 and 7 N lifts but not others). An example is provided by the participant represented by the solid red lines, who generated a relatively large lift height in one of the six first 2 N lifts. However, overall the spread of peak LF rates and maximum lift heights across blocks for a given participant appeared to be no greater in the coupled condition than in other conditions. To assess this issue quantitatively, we computed, for each participant and condition, the standard deviation of the peak LF rates and maximum lift height in first 7 N and first 2 N lifts, respectively. Repeated measures ANOVA failed to revealed significant differences in the SD of peak LF rates among the four conditions ($F_{3, 24} = 2.05; p = 0.133$). In contrast, a significant effect of condition was observed for the SD of maximum lift heights ($F_{3, 21} = 3.09; p = 0.049$). Corrected pairwise comparisons revealed that a reliable difference in the
SD of maximum lift heights between the informed and solo2 conditions \((p = 0.014)\); however, no other differences between pairs of conditions were observed \((p > 0.018\) in all 5 cases).

Figure 2.3 Distributions of initial peak LF rates and maximum lift heights. The left panels show cumulative distributions of peak LF rate for the first 7 N lifts and the right panels show cumulative distributions of maximum lift height for the first 2 N lifts. Separate distributions are shown for each participant in each condition. Each participant is represented by a consistent line color and type in all plots. The gray vertical lines indicate the mean value, average across all trials and participants, in each condition.
2.4.1.1 Load force scaling when lifting the 7 N object

Figure 2.4A shows, for all four conditions, mean peak LF rate, averaged across participants, for the last 2 N lift and the first, second, and last 7 N lifts. For the last 2 N lifts, corrected pairwise comparisons failed to reveal any significant differences between conditions ($p \geq 0.32$ for all 6 comparisons). For the first 7 N lifts, corrected pairwise comparisons revealed significant differences between all four conditions with the exception of the solo1 and solo2 conditions (solo1 v. coupled, $p = 0.025$; solo1 v. informed, $p = 0.004$; solo1 v. solo2, $p = 0.73$; coupled v. informed, $p = 0.020$; coupled v. solo2, $p = 0.006$; informed v. solo2, $p = 0.002$). This finding indicates that participants in the coupled condition used visual cues, obtained by watching the actor’s first lift of the 7 N weight, to scale LF predictively when subsequently lifting the object. However, this scaling was not as strong as in the informed condition in which participants were told that the weight was heavy. Similar corrected pairwise comparisons performed for the second 7 N lifts and for the last 7 N lifts failed to show significant differences between conditions ($p \geq 0.71$ for all 12 comparisons).
We also examined differences in initial peak LF rate between the four different lift types for each condition. For the coupled condition, corrected pairwise comparisons revealed significant differences between all conditions ($p \leq 0.015$) with one exception; no difference between the second and last 7 N lifts.

Figure 2.4 Average LF results. **A**: Mean initial peak LF rate, averaged across participants, for the last 2 N lifts ($L_{2N}$) and the first (1), second (2) and last (L) 7 N lifts in each condition. **B**: Mean initial peak LF rate for the first 7 N lifts as a function of block number and condition. **C**: Mean initial peak LF rates for the actor, averaged across sessions, for the four lift types in the coupled condition. **D**: Mean load phase duration, averaged across sessions, for the four lift types performed by the actor in the coupled condition. **A-D**: Error bars represent 1 SE based on participant means.
was seen ($p = 0.20$). For the informed condition, significant differences were observed between all pairs of conditions ($p \leq 0.015$) with two exceptions; there was no difference between the first and second 7 N lifts ($p = 0.18$) or between the second and last 7 N lifts ($p = 0.19$). For the solo1 condition, there was no difference between the last 2 N lifts and the first 7 N lifts ($p = 0.22$) but all pairwise comparisons were significant ($p < 0.003$ in all 5 cases). For the solo2 condition, the peak LF rate was slightly but significantly greater in the first 7 N lifts than in the last 2 N lifts ($p = 0.007$) and all other pairwise comparisons were also significant ($p \leq 0.011$).

Note that an increase in peak LF rate from the last 2 N lift to the first 7 N lift was also seen in lifts performed by the actor in the coupled condition (Figure 2.4C, see below). It is important to appreciate that this increase does not imply that participants, or the actor, anticipated the increase in weight. Although the initial peak LF rate typically occurs before lift-off when the weight of the object can be well predicted from previous lifts, in some trials the initial peak can occur around or just after the time of lift-off. Because lift-off leads to an abrupt cessation of LF increase (Johansson and Westling 1988), these later peaks will be smaller than they would be in trials in which the weight is unexpectedly heavy (such that lift-off does not occur). Importantly, lift-off does not lead to an abrupt cessation of grip force increase (Johansson and Westling 1988). Therefore, a key test of whether participants (in the solo conditions) or the actor (in the coupled condition) anticipated the increase in weight is whether peak GF rate increased from the last 2 N lift to the first 7 N lift. As we will show below, no such increase was observed, indicating that participants and the actor did not anticipate the switch from the 2 N to the 7 N weight.

Figure 2.4B shows, for each condition, mean peak LF rate, averaged across participants, for the first 7 N lifts in each of the six consecutive blocks of lifts performed with 7 N weight. Two-way repeated measures ANOVA revealed a significant effect of condition ($F_{3, 24} = 19.8; p < 0.001$) but no effect of block number ($F_{5, 40} = 1.34; p = 0.268$) and no interaction between condition and block number ($F_{15, 125} = 0.87; p = 0.599$). These results indicate that, within each condition, performance was stable across the session. Accordingly, the beneficial effect of action observation on force scaling in the coupled condition was present the first time the participant lifted the 7 N weight.
We also analyzed the behavior of the actor because we were interested in which cues the participants might have used to obtain information about the object weight from observing the actor. Figure 2.4C shows the actor’s mean peak LF rate for the last 2 N lift and the first, second, and last 7 N lifts in the coupled condition based on the average of means computed for each experimental session (i.e., participant). As expected, the peak LF rate in the first 7 N lift was similar to that observed in the participants’ solo conditions (figure 2.4A). Corrected pairwise comparisons indicated that, for both the first 7 N lift and the last 2 N lift, peak LF rate was smaller than for either the second or last 7 N lifts ($p \leq 0.001$ in all four cases), but did not differ significantly between the latter lift types. In addition, the peak LF rate was slightly but significantly greater in the first 7 N lift than in the last 2 N lift ($p = 0.004$). As noted above, this small increase in peak LF rate can arise because of biomechanical factors. These results indicate that the actor predicted a 2 N weight on the first 7 N lift but updated the weight prediction efficiently after a single lift.

One obvious cue about object weight that participants could have obtained by observing the actor is the duration of the load phase; i.e., the period during which LF is increased prior to lift-off. Prolongation of the load phase typically occurs in trials when the lifter underestimates object weight because the increase in LF does not result in lift-off and additional increases in LF are required to achieve lift-off (Johansson and Westling 1988; c.f. grey curves in figure 2.1D). Figure 2.4D shows mean load phase duration, averaged across sessions, for the actor’s last 2 N lifts and the first, second and last 7 N lifts in the coupled condition. Corrected pairwise comparisons between the four lift types failed to indicate significant difference in load phase duration between the second and last 7 N lifts. However, all other comparisons were significant ($p < 0.001$). The mean load phase duration for the first 7 N lifts was approximately 150 ms longer than for the last 2 N lifts. Likewise, the first 7 N lifts had longer load phase than the subsequent lifts in the 7 N blocks.
2.4.1.2 Grip force scaling when lifting the 7 N object

Although we were primarily interested in adaptation of load forces through action observation, for completeness we also examined the adaptation of grip forces. Figure 2.5 shows, for all four conditions, mean peak GF rate, averaged across participants, for the last 2N lift and the first, second, and last 7 N lifts. For the first 7 N lifts, corrected pairwise comparisons revealed a marginally significant difference between the solo1 and coupled conditions ($p = 0.054$) and a significant difference between the solo2 and the coupled condition ($p = 0.023$). Significant differences between each of the solo conditions and the informed conditions were also observed (solo1 v. informed, $p = 0.024$; solo2 v. informed, $p = 0.017$). However, no differences were seen between the two solo conditions ($p = 0.18$) or between the coupled and informed conditions ($p = 0.10$). These findings are consistent with the idea that participants in the coupled condition used visual cues, obtained by watching the actor’s first lift of the 7 N weight, to scale their fingertip force when subsequently lifting. Similar corrected pairwise comparisons performed separately for the last 2 N lifts, the second 7 N lifts, and for the last 7 N lifts failed to show significant effects between conditions ($p \geq 0.72$ for all 18 comparisons).

Figure 2.5 Average GF results. A: Mean initial peak GF rate, averaged across participants, for the last 2 N lifts (L$_{2N}$) and the first (1), second (2) and last (L) 7 N lifts in each condition. Error bars represent 1 SE based on participant means. B: Mean initial peak GF rates for the actor, averaged across sessions, for the four lift types in the coupled condition.
We also examined differences in peak GF rate between the four different lift types for each condition. For both the coupled and informed conditions, corrected pairwise comparisons revealed a significant difference between the last 2 N lifts and the three 7 N lifts ($p \leq 0.029$ in all six cases) but no differences between the three 7 N lifts ($p \geq 0.38$ in all six cases). For both of the solo conditions, corrected pairwise comparisons revealed that peak GF rate in both the last 2 N lifts and the first 7 N lifts were significantly smaller than in both the second and last 7 N lifts ($p \leq 0.002$ in all eight cases). However, in both the solo1 and solo2 conditions, there was no difference between the last 2 N lifts and the first 7 N lifts or between the second and last 7 N lifts ($p \geq 0.17$ in all 4 cases). The fact that peak GF rate did not increase from the last 2 N lift to the first 7 N lift in the two solo conditions indicates that participants did not anticipate the change in weight. The finding that peak GF rate significantly increased from the last 2 N lift to the first 7 N lift in the coupled condition (as it did in the informed condition) further indicates that participants benefitted from observing the actor.

Figure 2.5B shows the actor’s mean peak GF rate for the last 2 N lifts and the first, second, and last 7 N lifts in the coupled condition based on the average of means computed for each experimental session (i.e., participant). As expected, the peak GF rate in the first 7 N lifts was similar to that observed in the participants’ solo conditions (figure 2.5A). Corrected pairwise comparisons indicated that, for both the first 7 N lifts and the last 2 N lifts, peak LF rate was smaller than for either the second or last 7 N lifts ($p \leq 0.001$ in all four cases). However, no significant differences were observed between the first 7 N lifts and the last 2 N lifts ($p = 0.90$) or between the second and last 7 N lifts ($p = 0.48$). These results indicate that the actor predicted a 2 N weight on the first 7 N lift but updated the weight prediction efficiently after a single lift.

2.4.1.3 Maximum lift height when lifting the 2 N object

As outlined above (see Methods), using peak LF rate to assess load force scaling on the first 2 N lifts following blocks of 7 N lifts is problematic. However, we can indirectly assess load force scaling by examining the height of the lift. Figure 2.6A shows, for all four conditions, mean maximum lift height,
averaged across participants, for the last 7 N lift and the first, second, and last 2 N lifts. For the last 7 N
lifts, corrected pairwise comparisons failed to reveal any significant differences between conditions ($p \geq
0.13$ for all 6 comparisons). For the first 2 N lifts, corrected pairwise comparisons revealed that each solo
condition was significantly different than both the coupled and informed conditions (solo1 v. coupled, $p =
0.011$; solo1 v. informed, $p = 0.005$; solo2 v. coupled, $p = 0.004$; solo2 v. informed, $p = 0.001$). However,
no differences were seen between the two solo conditions ($p > 2$, note that $p$ values adjusted by the Holm-
Bonferroni test can exceed 1) or between the coupled and informed conditions ($p = 0.53$). These finding
indicate that participants in the coupled condition used visual cues, obtained by watching the actor’s first
lift of the 2 N weight, to scale LF predictively when subsequently lifting. This predictive scaling in the
coupled condition appears to be as strong as in the informed condition in which participants were told that
the weight was light. Similar corrected pairwise comparisons performed for the second 2 N lifts and for
the last 2 N lifts failed to show significant effects between conditions ($p \geq 0.13$ for all 12 comparisons).

We also examined differences in maximum lift height between the last 7 N lift and the first, second,
and last 2 N lifts for each condition. For the coupled condition, corrected pairwise comparisons failed to
reveal any differences between lift types ($p \geq 0.064$ in all 6 cases). For the informed condition, the
maximum lift height was greater for the first 2 N lifts than for the second 2 N lifts ($p = 0.020$), but no
other differences were seen ($p \geq 0.22$ in all 5 cases). For both solo conditions, the maximum lift height
was greater for the first 2 N lifts than for the three other lifts ($p < 0.001$ in all 6 cases), but no other
significant differences were observed ($p \geq 0.08$ in all 6 cases).
Figure 2.6 Average lift height results. **A**: Mean maximum lift height, averaged across participants, for the last 7 N lifts (L7N) and the first (1), second (2) and last (L) 2 N lifts in each condition. **B**: Mean maximum lift height for the first 2 N lifts as a function of block number and condition. **C**: The actor’s mean initial peak LF rates for the actor, averaged across sessions, for the four lift types in the coupled condition. **D**: Mean load phase duration, averaged across sessions, for the four lift types performed by the Actor in the coupled condition. **A-D**: Error bars represent 1 SE based on participant means.

Figure 2.6B shows, for each condition, maximum lift height, averaged across participants, for the first 2 N lifts in each of the six consecutive blocks of lifts performed with the 7 N weight. Two-way repeated
measures ANOVA revealed a significant effect of condition \( (F_{3, 21} = 22.4; p < 0.001) \) but no effect of block number \( (F_{5, 35} = 1.68; p = 0.166) \) and no interaction between condition and block number \( (F_{15, 105} = 1.16; p = 0.311) \). These results indicate that, within each condition, performance was stable across the session. Accordingly, the beneficial effect of action observation on force scaling in the coupled condition was present the first time the participant lifted the 2 N weight.

Figure 2.6C shows the actor’s mean maximum lift height for the last 7 N lifts and for the first, second, and last 2 N lifts in the coupled condition based on the average of means computed for each experimental session (i.e., participant). Corrected pairwise comparisons indicated that maximum lift height was greater for the first 2 N lift than the last 7 N lift \( (p = 0.001) \), the second 2 N lift \( (p = 0.002) \) and the last 2 N lift \( (p = 0.002) \) but did not differ significantly between the latter lift types. These results indicate that the actor predicted a 7 N weight on the first 2 N lift but updated the weight prediction efficiently after a single lift.

The results shown in Figure 2.6C indicate that participants could have obtained information about the weight of the object lifted by actor from the lift height, which was over 50 % larger on the actor’s first 2 N lift than on the actor’s previous 7 N lift. We would also expect the duration of the load phase of the actor’s lift to decrease on first 2 N lifts since the object would lift off sooner than expected. Figure 2.6D shows the actor’s mean load phase duration, averaged across sessions, for the last 7 N lift and for the first, second, and last 2 N lifts in the coupled condition. Corrected pairwise comparisons between the four lift types revealed that load phase duration was greater for the last 7 N lift than for all 2 N lifts \( (p < .001 \) in all 3 cases), that load phase duration was smaller for the first 2 N lift than the second \( (p = 0.023) \) and last \( (p = 0.020) \) 2 N lifts, and that there was not difference between the second and last 2 N lifts. The mean load phase duration for the first 2 N lifts was approximately 100 ms longer than for the last 7 N lifts, and therefore load force duration may have been an effective cue indicating that the weight lifted by the actor had decreased.

In summary, our results indicate that in the coupled condition, participants gained information from observing the actor’s lifts and used this information to adapt their motor output when the weight of the object changed. Overall, this adaptation was almost as strong as when participants were verbally informed...
about the weight change. The adaptation seen in the coupled condition was consistent throughout the experimental session, including the first lift of the 7 N weight following the initial block of lifts of the 2 N weight and the first lift of the 2 N weight following the initial block of lifts of the 7 N weight.

2.5 Discussion

Our results clearly support our hypothesis that participants extract information about object weight when observing another person lift an object and that they make use of this information when they subsequently interact with the object. Specifically, we demonstrate that when lifting weights previously lifted by another person, people naturally (i.e., without explicit instruction) extract information about weight. Moreover, they use this information effectively when scaling lifting forces. Therefore, this result extends previous work showing that people can make relative judgments about weights lifted by others (e.g., Runeson and Frykholm 1981; Bingham 1987; Shim and Carlton 1997; Hamilton et al. 2007).

Our results are broadly consistent with those reported by Meulenbroek and colleagues (2007) but show a stronger effect of action observation on lifting performance. These authors examined pairs of participants performing trials in which one participant (corresponding to the actor in our study) lifted an object and placed it into a shared workspace and the second participant then lifted and retrieved the object. They determined, for both placers and retrievers, the change in lift height, relative to the previous trial, in trials in which the weight was unexpectedly changed. When the weight was unexpectedly light, a marked increase in lift height was seen in both placers and retrievers, but the increase was slightly smaller in retrievers. In contrast, we found what appears to be a much stronger benefit of action observation. Specifically, in first 2 N lifts in the coupled condition, we found that whereas the actor’s lift height increased substantially (as expected), participants’ lift heights were not reliably greater than their lift heights in their previous lifts (i.e., the last 7 N lifts). Meulenbroek and colleagues (2007) also reported a modest benefit of action observation in trials in which the object was unexpectedly heavy. However, the latter result is somewhat unclear because, in the previous trial, lift height was greater in placers than retrievers. Moreover, lift height is not an ideal measure of expected weight in trials in which weight is
unexpectedly increased because it will be affected by corrective actions taken when the object does not
lift off at the expected time (Johansson and Westling 1988). In the current study, we used the initial peak
rate of change of load force, which occurs prior to lift off, to assess expected weight in trials in which
weight was unexpectedly increased. Again, we found a very robust effect of action observation. Specifically, in first 7 N lifts in the coupled condition, participants (unlike the actor) exhibited a marked
increase in peak load force rate that was almost as strong as when participants were verbally informed
about the weight change. Several factors may have contributed to the stronger effects of action
observation seen in the current study in comparison to the previous study by Meulenbroek and colleagues
(2007). In the previous study, four different objects were used, which varied in both weight (230 or 835 g)
and size (25 cm high cylinders with diameters of 2.5 or 6.5 cm), and the cylinder was changed every 3
trials. Thus, the participants presumably knew when a weight change might occur. Moreover, as the
authors showed, size affected lift height independently of weight.

Our findings complement work by Mattar and Gribble (2005) demonstrating that adaptation of point-
to-point movements of a hand-held object that exerts a novel and unusual load on the hand can be
enhanced (or impaired) by first observing an actor performing the task under the same (or opposite) load
conditions. These authors found that action observation had a small but significant effect on initial
performance, but that hundreds of trials were still required for full adaptation (as in the control condition
without action observation). In contrast, we found that action observation had a dramatic effect on initial
performance such that performance was close to being completely adapted on the first lift in the coupled
condition. This difference can be expected from the two-component model of skill learning outlined in the
Introduction. In the Mattar and Gribble (2005) study, full adaptation would have required observers to
learn both the structure and parameters of the load because both the equations of motion of the load and
the parameters of these equations were novel (Braun et al. 2009; Wolpert and Flanagan 2010). However,
in the current study only parametric learning was required because the participants were familiar with the
structure (i.e., equations of motion) of the load. That is, our participants only had to adapt, via action
observation, their force output to changes in object weight (or mass). Note that by using the term
parametric learning, we do not mean that observers learn the precise weight parameters. Rather, we are referring to the fact that they gain information about a parameter (i.e., weight) of an object with familiar dynamics.

When lifting objects, people scale their fingertip forces to the expected weight of the object and also generate a prediction about when they will receive sensory events signaling lift-off, including discrete tactile signals from mechanoreceptors in the hand that are sensitive to mechanical transient events (Westling and Johansson 1987). When an object is heavier or lighter than expected, sensory events signaling lift-off do not occur at the predicted time and the resulting mismatch between predicted and actual sensory events triggers task-protective corrective actions. For example, when the object is heavier than expected, the absence of the predicted sensory events signaling lift-off triggers a corrective action that involves probing increases in vertical load force (Johansson and Westling 1988; Wolpert and Flanagan 2001). This mismatch between predicted and actual sensory events also leads to an updating of memory related to object weight, which improves future motor output and sensory predictions (Flanagan et al. 2006; Johansson and Flanagan 2009).

It has been suggested that such error-based learning, which is a critical component of motor learning through practice (Wolpert et al. 2001; Shadmehr et al. 2010), may also underlie motor learning through observation (Wolpert et al. 2003; Mattar and Gribble 2005; Oztop et al. 2005; Brown et al. 2010). In a follow-up to the Mattar and Gribble (2005) study, Brown and colleagues (2010) showed that motor learning via observation is best when observers view movements with large errors. Based on the current results, we suggest that when watching another person lift an object, observers update information about the object’s weight by predicting when the object will lift off and comparing this time to the viewed lift off time. Of course, participants may also predict and evaluate other aspects of the actor’s lift such as lift speed and height. For example, when an object is lighter than expected, the height of the lift will typically increase and the earlier-than-expected lift-off triggers a corrective action that brings the object back to the intended height (Johansson and Westling 1988). Moreover, observers can use other cues, such as hand shape, to estimate the weight of objects lifted by others (Alaerts et al. 2010b). In general, actors might
also provide other facial or even verbal cues about unexpected weight changes when lifting objects. However, the actor used in the current study was instructed not to make any facial expressions and to act in a consistent manner throughout each experimental session. Moreover, previous work has shown that when watching another person perform an object manipulation task, observers direct their gaze at the objects as they are grasped and lifted and rarely look elsewhere (Flanagan and Johansson 2003; Falck-Ytter et al. 2006; Rotman et al. 2006; Webb et al. 2010).

One way in which observers may generate predictions about lift performance, including lift-off time, would be to covertly simulate the observed action in approximate synchrony with the actor (Iacoboni et al. 1999; Iacoboni et al. 2001; Rizzolatti et al. 2001; Wolpert et al. 2003; Rizzolatti and Craighero 2004; Rizzolatti and Fabbri-Destro 2008). Behavioural evidence in support of this possibility comes from studies of observers’ gaze behaviour when watching object manipulation tasks. When watching an actor perform a block-stacking task, observers’ gaze fixations closely resemble those of the actor in both space and time (Flanagan and Johansson 2003; Rotman et al. 2006). Specifically, observers, like actors, proactively fixate blocks that the actor is reaching for in order to grasp, and locations where the actor is reaching, with the block in hand, to place the block. It has been argued that, in action, task specific eye movements are called by the motor plan such that they provide task critical visual (and proprioceptive) information at the appropriate times (Land and Furneaux 1997). If observers run a covert sensorimotor plan when watching action, then task specific eye movements that are similar to the actor’s would be expected.

Support for the idea that observers of action activate sensorimotor representations is also provided by neurophysiological studies showing that sensorimotor areas and circuits engaged when performing an action task are also recruited when observing the task (Rizzolatti et al. 2001; Rizzolatti and Craighero 2004; Malfait et al. 2010). Of particular relevance to the current work are recent results based on transcranial magnetic stimulation indicating that motor cortex facilitation in observers is specific to the muscles used by an actor lifting objects and scales with the force applied to the objects, i.e., to object weight (Alaerts et al. 2010a; Alaerts et al. 2010b). However, whether action simulation is used to generate
predictions about observed action remains a matter of active debate (e.g., Brass et al. 2007; Aglioti et al. 2008; Hesse et al. 2009).

Because participants only lifted two weights (i.e., the 2 and 7 N weights), it seems likely that, in the informed condition, they learned to use the verbal information about the change in weight to access sensorimotor memory of these two weights. The fact that their lifting performance on the very first weight change in the informed condition was quite accurate suggests they may have assumed that the weights would be the same as in the prior conditions. In principle, in the coupled condition, participants could have also remembered the two weights, or the lift forces required to lift them, and then used visual information from the actor's lift in the coupled condition to select the appropriate memory. Alternatively, in the coupled condition participants could have directly estimated object weight (or the change in object weight) by simulating the actor’s lifts and comparing predicted and actual performance parameters (e.g., lift height and load phase duration). One argument against the former possibility is that, whereas participants anticipated weight changes in the coupled condition, they continued to adapt load force in subsequent lifts, indicating that they did not fully adapt load force on the first lift of the new weight. However, it is also possible that participants were simply being conservative in terms of changing their force output given uncertainly about the weight.

In summary, our results indicate that people naturally encode information related to object weight when watching another person lift objects. That is, in the coupled condition participants exploited visual information from the actor’s lifts to adapt their motor output to the weight of the object to be lifted despite the fact that they were not informed about the structure of the experiment (i.e., the fact that they would lift the same sequence of weights as the actor). Moreover, this adaptation occurred right from the start. That is, in the coupled condition participants effectively adapted their force output the first time they watched the actor lift the 7 N weight after the initial block of lifts of the 2 N weight and the first time they watched the actor lift the 2 N weight after the first block of lifts of the 7 N weight. The efficacy of this predictive adaptation is presumably context-specific and may depend on the likelihood that the observer will be required to lift the object as well as on their level of engagement and attention. In our experiments, we
used a single object and two weights and this may have facilitated parameter learning through action observation. In future work, we plan to investigate such learning in situations with multiple weights and multiple objects.

2.6 References


Alaerts K, Senot P, Swinnen SP, Craighero L, Wenderoth N, Fadiga L (2010a) Force requirements of observed object lifting are encoded by the observer's motor system: a TMS study. The European journal of neuroscience 31: 1144-1153

Alaerts K, Swinnen SP, Wenderoth N (2010b) Observing how others lift light or heavy objects: which visual cues mediate the encoding of muscular force in the primary motor cortex? Neuropsychologia 48: 2082-2090


Chapter 3

Observation to Objects Relevant to the Observer's own Future Actions

3.1 Abstract

Observers, like actors, closely monitor object-oriented actions by shifting their gaze to object contact events. Here we test the hypothesis that observers selectively attend to objects relevant to guiding their own future actions. In our task, a demonstrator lifted a central start object with both hands and then simultaneously lifted two target objects, located on either side of the start object, with both hands. The participant was then cued to lift one of the two target objects. We occasionally increased the weight of one of the target objects. In experiment 1, we repeatedly cued each target object in blocks of trials making the subsequent target object largely predictable to participants. In experiment 2, the cued object was randomized but we exclusively changed the weight of one particular target object in a given block of trials making this object more challenging for the participants to lift. We recorded participants' eye movements as well as their fingertip lifting forces in order to test the hypothesis that observers preferentially attend to object contact events relevant for guiding their own future actions. Participants engaged in highly variable gaze behavior when observing the demonstrator lift the target objects bimanually. We report a strong effect of cueing on gaze, with participants clearly favoring the previously cued object, but no apparent effect of variable versus constant weight schedules. We also report robust social learning of object weight, in particular in experiment 1, but did not find a clear relationship between gaze strategies and subsequent lift performance. Our results indicate that participants are able to identify increases in the weight of one of the two simultaneously lifted objects even when not fixating that object. While the relationship to selective social learning turned out to be more complex than we had anticipated, overall we demonstrate a strong influence of the observer's own action goals in action observation.
3.2 Introduction

When looking at other people engaged in activities, observers, like actors, closely monitor object-oriented actions with their gaze being drawn to object contact events in particular (Flanagan & Johansson, 2003). However, observers frequently encounter multiple concurrent actions performed either by multiple actors or one actor engaging with two objects bi-manually. How do observers monitor multiple actions which compete for their attention?

We have previously emphasized the observer's motives in action observation. The actor's gaze serves both to guide their ongoing actions and monitor fulfillment of their action goals (Johansson, Westling, Bäckström, & Flanagan, 2001; Land, 2006; Triesch, Ballard, Hayhoe, & Sullivan, 2003), and we have proposed that gaze in action observation is likewise driven by the observer's goals – which do not necessarily coincide with those of the actor (Flanagan, Rotman, Reichelt, & Johansson, 2013). Observers likely have a range of concerns and interests and the typical emphasis on the function of action observation as extracting the actor's goals thus provides only a partial account of action observation (e.g. Rizzolatti & Sinigaglia, 2010). In particular, observers who are already familiar with the goals and sequence of upcoming actions when observing routine and repetitive activities are still drawn to monitor the unfolding action sequence, and actually use their prior knowledge to shift their gaze to object contact events much earlier than they could for unpredictable actions (Flanagan & Johansson, 2003; Rotman, Troje, Johansson, & Flanagan, 2006; Henrichs, Elsner, Elsner, Wilkinson, & Gredebäck, 2014).

Here we investigate the proposal that observers - like actors - monitor object manipulation (in particular contact events) in order to update their estimates of objects to facilitate guiding their own future actions. In an object lifting task participants first observed a demonstrator lift a start object and then lift two target objects with both hands simultaneously, creating competition for gaze, after which participants were cued to lift one of the target objects themselves. In random trials, the weight of one of the target objects was increased to heavy by means of a small switchable electromagnet. We varied the probability of a target object being cued, and the probability of a target object changing weight, respectively, to test
the prediction that observers will 1) preferentially attend to actions on objects they deem more relevant to their own actions, and 2) preferentially attend to actions on objects whose properties they are uncertain of.

To our knowledge this is the first study investigating action observation in detail in a controlled setting where observers can directly make use of social learning for guiding their own actions. A number of studies of action observation have focused on how observers' motor skills and visual experience respectively modulate action recognition and motor resonance in parieto-frontal cortical circuits (Aglioti, Cesari, Romani, & Urgesi, 2008; Casile & Giese, 2006; Gardner, Goulden, & Cross, 2015). Fewer studies have examined the effects of visual familiarity and skill practice on eye movements (Green, Li, Lockman, & Gredebäck, submitted; Möller, Zimmer, & Aschersleben, 2014). However, these studies did not describe social observational learning in terms of benefits to the observers' own subsequent action performance or investigate how an ability to actually apply what is learned from observing itself affects gaze in action observation and vice versa.

From a motor control perspective, a series of studies have documented motor learning from observing for learning novel reach dynamics (Mattar & Gribble, 2005), with observation of another's reach errors activating cortical networks which overlap in part with those involved in processing self-generated movement errors (Malfait et al., 2010). Rapid learning of object properties such as change of an object's mass have been described for object lifting (Reichelt, Ash, Baugh, Johansson, & Flanagan, 2013), and observing lifting of heavy vs light objects has been shown to modulate the excitability of primary motor cortex, already in anticipation of the actor's lift (Alaerts, de Beukelaar, Swinnen, & Wenderoth, 2012). In turn, disrupting activity in motor cortex in observers using TMS over the hand areas has been shown to delay object directed gaze shifts when observing point-light hand actions (Elsner, D’Ausilio, Gredebäck, Falck-Ytter, & Fadiga, 2013). However, these studies have looked at very different model systems and either did not record gaze or in the case of the latter paper did not provide a detailed analysis of gaze strategies.
3.3 Methods

3.3.1 Participants

One demonstrator (woman, 23 years) and 14 participants (10 women, 19-24 years) took part in experiment 1, the same demonstrator and 11 participants (8 women, 19-23 years) completed experiment 2. One participant was removed from the dataset of experiment 2 due to recurring calibration problems with eye-tracking. All participants gave informed consent and were uninformed about the research question. The local university ethics board approved the study, which complied with the Declaration of Helsinki.

3.3.2 Setup

Participants (seated) and the demonstrator (standing) repeatedly lifted objects either by themselves or taking turns (see experimental conditions), see figure 3.1A. Three objects of identical size were used (5 cm$^3$ hollow cubes made from the opaque polyoxymethylene plastic Delrin®) positioned along a line 19 cm apart from each other. The left and right target objects were both black while the color of the central start object was green. The target objects had a handle mounted on the top of the cube which was instrumented with 2 force–torque sensors (Nano 17 F/T, ATI Industrial Automation, Garner, North Carolina) that measured the forces applied by the tips of the thumb and index finger. Flat circular disks, covered by medium-grain sandpaper, capped each sensor, oriented in parallel vertical planes, separated by a distance of 4 cm (shown in figure 3.1B). The start object had a similar handle but was not instrumented. A miniature electromagnetic position sensor (Polhemus Liberty, Burlington, Vermont), attached to the top of each object, was used to measure the height of the lift. The weight of the target objects was set to either light (230 grams) or heavy (650 grams) by means of a small switchable electromagnet (see figure 3.1C), allowing for nearly instantaneous weight changes. All magnets were engaged (turned off, on, or on-and-off) at the beginning of every trial while the start tone was played and then were maintained in a constant
state throughout each trial so that possible minute clicking sounds made by weights attaching to the rods at no point gave cues about upcoming target object weight to participants or the demonstrator.

Figure 3.1 A Top view of the experimental setup showing the demonstrator and participant's positions and the position of the three objects used. B Object with handle instrumented with force sensors. C Schematic showing the switchable magnets for setting object weight in operation. In this example bimanual lift, the left magnet is engaged - turning the left target object heavy - while the right magnet is disengaged. The weight of the central start object was set to light throughout the experiment. Guides ensured that demonstrator and participant returned the object as well as the additional weight to a central position.

Finally we recorded pupil and corneal reflection position and pupil diameter using a head mounted infrared video-based eye tracker (RK-826PCI ISCAN, Inc., Burlington, MA) also instrumented with an electromagnetic position sensor to correct for participants' head position.
3.3.3 Data Analysis

Force and position signals were recorded at 250Hz and eye position was recorded at 240Hz. All signals were interpolated to 1KHz and smoothed using a fourth-order, zero-phase lag, low-pass Butterworth filter with a cutoff frequency of 14 Hz (forces and positions) and 25 Hz (eye position), respectively. Data collection was performed in LabVIEW (National Instruments) and data analysis in MATLAB (MathWorks). We computed participants' line of sight in 3 dimensions from the differential pupil and corneal reflection positions and participants' head position. Line of sight was then intersected with the object plane to compute at horizontal and vertical coordinates. For calibration, participants were instructed to rotate their head sideways and up and down while keeping their gaze fixed on a calibration point located in the center of the workspace (not visible during the experiment). For details on the 3D eye tracking algorithm and calibration method see Ronsse, White, & Lefèvre (2007). For analysis of gaze, we first identified and removed blinks based on the pupil diameter signal, and used regions of interest around each of the three objects (5 cm to left and right, 10 cm up and down from the object center) to categorize object-directed gaze both in terms of location (left, center, and right) as well as in terms of cue history (previously cued and previously uncued, cue switch trials). For gaze preference scores, gaze directed at the previously cued target object was weighed as (+1) per trial, while gaze at the previously uncued target object was weighed as (-1), with gaze at the start object counted as a neutral (0). Thus a participant who is looking back and forth between the two target objects symmetrically (for equal durations within the time window of interest) in a trial and participant who constantly fixates the central start object throughout a trial would both receive a neutral gaze preference score for that trial.

We added the forces in the vertical direction from the two sensors in each target object handle to compute the vertical force (vF) and calculated the rate of change of vertical force (vF Rate) with respect to time using a first-order central difference equation. We characterized each trial in terms of lift phases (start object lifts, bimanual lifts, and participant lifts) with reach phases in between. The beginning and end of the start object lift phase was identified when the central object (not instrumented with force
sensors) was first lifted higher than .25 cm and back below this level, while target object lift onset times were identified when the vertical force exceeded .2 N. We defined the beginning and end of demonstrator's bimanual lift phase as the force at either object exceeding this threshold and at both objects dropping below this threshold, respectively. We normalized the phase durations of each trial by scaling gaze position traces for each phase to the median phase duration which were used for visualization purposes only (see figures 3.2-3.5, 3.7). Gaze preference scores were calculated as an overall difference score within the bimanual lifting phase of each trial. To characterize the benefit of learning by observing on object lifting, we calculated the duration of the lift-off sub-phase (lift-off duration) for each target object lift defined as the time between lift start (vF > .2 N), and imminent lift-off (vF > 1.5N for light and vF > 6 N for heavy objects) to exclude fluctuations of vertical force involved in physical lift-off. Lift-off duration is delayed for unexpectedly heavy objects (Johansson & Westling, 1988) and constitutes a functional measure reflecting task achievement.

3.3.4 Procedure

*Experiment 1:* Participants performed one solo condition (for details see below) followed by 4 separate social conditions with short breaks between conditions. Each condition consisted of 50 trials. Participants were cued to lift a given target object in blocks of 10±2 trials and therefore came to expect to be cued to keep lifting the object their lifted in the previous trial. As the cue indicating which target object to lift was changed 4 times, this resulted in 5 blocks of repeated lifts of a target object and 4 cue switch trials in each condition. A graphical overview of the order of condition and composition of trials within a condition is given in table 3.1.
Experiment 2: Participants performed one solo condition, followed right away by either a social or a passive condition and a short break. This was repeated four times resulting in 4 solo, 2 social, and 2 passive conditions. Participants were cued to lift target objects at random, so were not able to anticipate which object they would have to lift later in social conditions. In each condition, either the left or the right target object would exclusively become heavy in 16 of the 50 trials while the other one would always be light. The location (left or right) of the object that was to heavy in a subset of trials was counter-balanced over conditions but was held constant for each solo condition and the subsequent social or passive condition. Participants were cued to lift the heavy object in 8 trials of those 16 trials, see table 3.2. This allowed participants to potentially anticipate which target object weight was variable and which stayed constant, respectively.

Table 3.1: Order of conditions and trial composition of experiment 1. Note that there were 4 cue change trials in each condition (shown in red). In 16 of the 50 trials (randomly interspersed) one object was set to heavy weight. The participant was cued to lift the heavy object in 8 trials (heavy lift trials), and was cued to lift the light object in the remaining 8 trials. One of the 8 heavy lift trials was also a cue switch trial, resulting in one lift per condition where the participant was cued to lift the other, previously not cued object, and that object was set to heavy (heavy switch trial). Note that as participants first observed the demonstrator lift before receiving their cue.
In **solo conditions**, participants grasped and lifted one of the target objects in each trial after being instructed by a tone which object to lift (left or right). After 750ms from lift onset a neutral tone was played instructing them to replace the object.

In **social conditions** the demonstrator, prompted by a distinct auditory cue, first grasped, lifted, and replaced the central start object with both her hands and then moved her arms outward to grasp, lift, and replace both target objects at the same time using both her hands. After the demonstrator concluded her lifts, a tone was played instructing the participant to grasp, lift, and replace either the left or the right

---

**Table 3.2**: Order of conditions and trial composition of experiment 2. All participants completed 4 solo conditions, and 2 social and 2 passive conditions in counterbalanced order. A and B refer to the object location (either left and right, counterbalanced between participants). Participants first completed a solo condition followed by either a social condition or a passive condition with the object of variable weight located on a given location. Next they completed another solo condition followed by the remaining condition type (social or passive) which they did not yet complete at the other location. The double-arrows indicate counterbalancing order. In the second half of the experiment, participants again completed two solo conditions followed by a social and a passive condition (again with the order counterbalanced over participants) with the object of variable weight now set to the other location. Objects were randomly cued to be lifted and of the 16 trials where one object was set to heavy (always at the same location), participants were cued to lift the heavy object in 8 trials (heavy lift trials).
target object. Participants always lifted with their preferred hand (their right hand except for one participant in experiment 2).

In passive conditions (experiment 2 only), the experimenter lifted and replaced first the start object and then both of the target objects simultaneously, and then repeated this sequence without the participant performing any manual actions.

The central start object was included to make the timing of the demonstrator's target object lifts predictable to observers as well as to provide a neutral starting position for the participants' gaze. For 34 out of the 50 trials all object weights were set to light in all conditions. For the remaining 16 trials (about one in three) the weight of either one of the target objects was set to heavy. In half of those 16 trials this object was then cued to be lifted by the participant, and in the other half the participant was cued to lift the other, light object, resulting in 8 heavy weight trials for the participant per condition. Heavy weight trials were randomly interspersed, but did not repeat for the participant. Object weights were kept constant within each trial, so that in social conditions participants could reliably extract information about the target object weights from observing the demonstrator's lifting action.

Practice trials and instructions: In both experiments, participants received about 40 practice trials of individual lifts similar to a solo condition to familiarize themselves with lifting the objects (light and several unpredictably heavy lifts), and the auditory cues indicating lifting of the left or right target object, respectively. Participants were instructed to grasp the object at the handle using a precision grip, confidently lift it vertically to a height of about 2cm, hold the object stationary, and then replace it on hearing the neutral tone. At the start of the first social condition, participants were told that they were now performing the experiment together with the demonstrator, taking turns, and that they would get to see the demonstrator lift first. Throughout the experiment we were relying on participants' being intuitively inclined towards attending to object actions and making smooth and controlled lifting movements. We did not instruct participants to attend to the lifts of the demonstrator, nor did we give them feedback regarding their lift performance once the initial practice session was concluded.
The same demonstrator, naïve to the purpose of the experiment, was used throughout. The demonstrator was instructed to always lift with the expectation of a light object and received extensive practice to ensure consistent timing of the lifting sequence throughout the experiment.

In debriefing we asked participants whether watching the demonstrator lift first had benefited their own lifting and in experiment 2 they were also asked whether they had noticed a pattern in the distribution of heavy weights (which was always at one side only in each block of conditions).

3.4 Results

3.4.1 Experiment 1

Participants almost always directed their gaze at one of the three objects and in almost all trials monitored the demonstrator lifting the central start object. Gaze behavior became strikingly variable beginning with the reach phase in preparation for the bimanual lift, with individual participants pursuing highly different gaze strategies. Therefore, in the absence of gaze behavior representative for most participants, we first show example gaze and lifting traces from 3 selected participants and analyze the diversity of gaze behavior within the population before moving to summary statistics.

3.4.1.1 Individual results

Figure 3.2A shows gaze behavior and lifting performance of participant P1. In all but two trials, P1 monitored the demonstrator lifting the central start object. In many trials, gaze is shifted to the previously cued object already while the demonstrator is reaching for the target objects. In a subset of trials gaze is kept at the start object, with a few gaze shifts between the objects during the demonstrator's bimanual lifts. In those trials where gaze is not already directed at the previously cued object, gaze is shifted there at cue onset before the participant's own lift (separation of blue and red traces). Therefore, throughout the bimanual lifting phase, gaze is directed preferentially at the previously cued target object – the object likely to be cued next – both for light object trials (dark trace in figure 3.2B), as well as heavy object trials (strong green and orange traces on figure 3.2C). This preference is already evident during the reaching
phase before the demonstrator lifted the target objects. In addition, when the previously uncued object was heavy - and thus did not lift off at the expected time – gaze was still much more likely to remain at the previously cued object and was drawn to the uncued heavy object in only a few trials (compare thick orange and green traces in figure 3.2C). Conversely, participant P1 showed no discernible bias to the left or right object (light traces in figure 3.2B-C) which were cued equally often overall. In a majority of light weight trials, P1 either exclusively looked at the cued object or shifted gaze from the neutral start object to the previously cued object within the bimanual lifting phase, looking towards the uncued object only very rarely (figure 3.2D-E). P1 showed a clear benefit from social learning, as shown by the steeper increase in vertical force when lifting a heavy object in the solo compared to the social condition, resulting in P1 lifting up a heavy weight about 200ms faster in social conditions (figure 3.2F-G).
Figure 3.2: caption see next page.
Figure 3.2: Gaze behavior and lifting performance of an example participant (P1) in experiment 1.  
A-C Gaze behavior in social conditions over normalized lift phases (gray background: demonstrator lifting the start object, demonstrator lifting both target objects simultaneously, the participant lifting the cued target object) and reach phases (white background). Note that the duration of the key bimanual lifting phase is extended in heavy weight trials, due to the additional time needed for the demonstrator to adjust to an unpredictable weight. A Horizontal gaze traces from all 134 light weight trials showing fixations on and gaze shifts between the start object (center) and target objects (right and left). B Dark gray trace shows the probability of Gaze (mean over all light trials) being directed at the previously cued target object (1), or at the previously uncued target object (-1), with gaze at the start target being counted as neutral (0). Light gray shows the side bias, computed similarly between the right (+1) and left target object (-1). C Probability of Gaze at the previously cued target object for all heavy trials (not including cue switch trials), shown separately for the heavy target object being previously cued (green trace) or uncued (orange trace), respectively. Also shown is the bias towards right (1) and left (-1) target objects (light green and light orange traces). D Proportion of trials where gaze first shifted to the cued or uncued target, respectively. Note that while participants always shifted their gaze to one of the target object eventually, they may do so only on hearing their auditory cue after the demonstrator has replaced the target objects. The remaining bar graphs show the probability of gaze (mean proportion of fixations) for each object in light weight trials for specific durations within the key bimanual lifting phase for the cued target object, central start object, or uncued target object. LiftOff: window of 500ms after the start of the demonstrator's bimanual lift. Overall: the complete bimanual lifting phase. Place: the final 500ms until the demonstrator has replaced the objects. E Bar graphs show the proportion of gaze sequences within the bimanual lifting phase for light weight trials grouped for sequence length. Particular gaze sequences starting at each object are denoted as stacked bars (i.e. trials where gaze started at the central object and then shifted towards the cued target object are shown as the gray-to-green bar in the second group). Gaze had to be localized within the region of interest for the central object for longer than 60ms to exclude hits from passover saccades between the target objects. F Vertical force (vF) and rate of change of vertical force (vF Rate) traces in heavy weight trials for the solo condition (purple, n = 8), previously cued trials (green, n = 28), and previously uncued trials (“heavy-switch”, orange, n = 4) from the social conditions. Dashed lines show the vertical force thresholds (.2N and 6N) used to calculate lift-off durations. G Mean lift-off durations for the three heavy weight trial types with standard errors.
Figure 3.3 Participant P7, description see caption figure 3.2.
Figure 3.4 Participant P8, description see figure 3.2.
Participant P7 (see figure 3.3) again almost always monitored the demonstrator lifting the central start object, and then focused on the target objects and rarely looked at the start object again (gray bars in figure 3.3D). A first preference towards the previously cued object is shown as the demonstrator is contacting and lifting up the target objects in the bimanual lifting phase. As the demonstrator is holding up the target objects, this preference recedes again and P7 is equally likely to attend to each target object. As the demonstrator is putting the target objects back down, P7 begins to develop a robust preference for the previously cued object. The small tendency of shifting gaze to the cued side first is overlaid with a partial side and sequence preference for looking at the left target object, switching to the right, and then back to the left (see light trace in figure 3.3B), regardless of which target object was being cued. Unlike the previous participant (P1), P7 showed a robust reaction to the demonstrator lifting a heavy weight with gaze being drawn to the side of the heavy target object almost without fail (compare strong green and orange traces in figure 3.3C). Overall P7 performed many more saccades between objects as compared to P1, typically rapidly shifting gaze between the target objects 3 or more times, e.g. from the cued to the uncued and back to the cued target object in more than a third of light weight trials (figure 3.3E). Like P1, P7 showed highly robust social learning, cutting lift duration of heavy objects in half for social conditions (a benefit of about -300ms) compared to the solo condition (figure 3.3F-G).

Finally, a very different gaze strategy was exhibited by participant P8 (figure 3.4). In about 75 percent of trials, gaze remained fixed on the central start object until the auditory cue was given. For the remaining quarter of trials, gaze is either directed at the cued object exclusively or in combination with the neutral start object, and almost never at the previously uncued target object (see figure 3.4D). Overall this results in a small but robust preference for the cued target object. This holds for heavy weight trials as well, as P8 does not show any discernible overall reaction to the demonstrator lifting a heavy target object (figure 3.4C). Note that of the two target objects, P8 attended to the cued target object first 90 percent of the time, but for this participant the choice between the target objects typically occurred late in a trial after the demonstrator had already replaced the target objects. This behavior renders the measure of the first saccade to a target object unsuitable for characterizing gaze behavior for social learning of object weight.
P8 showed a robust social learning benefit as well, quantified as a reduction in lifting duration of around 200ms (figure 4.4F-G).

3.4.1.2 Group results

At the group level, participants preferentially directed their gaze at the previously cued target object as shown in figure 3.5A, but differed markedly in terms of the extent of this preference, ranging from near exclusive focus on the previously cued object for participant P9 to no preference for participant P3. Some participants preferentially looked at the previously cued target object already in anticipation of the demonstrator lifting the target objects (P5, P9, P13, and P14), whereas other participants built up their preference over the course of the bimanual lifting phase (e.g. P6 and P12), while still others did not show preference until right before the cue was given (P3 and P11). As shown in figure 3.5B, a few participants showed a massive bias for a physical side, i.e. P6 and P12 tended to begin looking at the right and the left target object, respectively, before shifting their gaze to the start object or the other target object. While a few participants show a clear bias either to the left (e.g. P8) or to the right side (e.g. P6), respectively, especially at the beginning of the demonstrator's bimanual lift, no consistent bias is evident overall.
Figure 3.5 caption see next page
Figure 3.5: Gaze behavior and lifting performance from all 14 participants. A Color coded traces shows the probability of gaze (mean over all light trials) being directed at the previously cued target object (1), or at the previously uncued target object (-1). B Color coded traces show side bias, computed similarly as between right (+1) and Left target object (-1). C Mean difference scores for gaze probability for the previously cued minus uncued target object from all light trials, and D reduction in lift-off duration for social conditions compared to solo conditions shown, respectively, for all participants (color-coded scatterpoints) and group means and standard errors (blue bars). E Scatterplot of the relationship between these two measures with a linear regression line overlaid.

To quantify the preference for the previously cued versus the previously uncued target object, we computed the difference score between the differential proportion of gaze for the previously cued (+1) and uncued (-1) object over the duration of the bimanual lifting phase in light weight trials for each participant. Light weight trials provide a clear measure since the demonstrator showed symmetrical lifting performance, identical over all the included trials, so any systematic difference to gaze to the target objects is due to the gaze strategies of observers rather than in response to the demonstrator's lifting action. A one sample t-test revealed that participants as a group showed a preference towards the previously cued target significantly more often than expected by chance (mean difference score of 0.4, sd = 0.28, $p < 0.001$). Next we quantified the benefit of social learning for each participant defined as the difference in mean lift duration for heavy objects between the solo and the social conditions (shown in figure 3.5D). Most participants exhibited robust social learning of object weight with lift duration reductions but showed high variability between participants (20-300ms). The lack of a social learning benefit for object lifting in a small subset of participants has already been reported in a previous study with only one target object (Reichelt et al., 2013). As a group, participants' social learning benefit was significantly different from zero (one sample t-test, mean duration reduction 167ms, sd = 86.6, $p < 0.00001$).

Strictly speaking, part of this reduction in lifting duration could be due to a general practice effect as the solo condition was performed ahead of the social conditions in experiment 1. However, we have
previously shown the size of the practice effects to be far smaller than the benefit due to social learning (Reichelt et al., 2013), and in experiment 2 participants did not show any decrease in lifting duration in the course of the four solo condition they performed (group mean lift-off duration of 426ms, 416ms, 469ms, and 439ms, respectively, for solo condition in the order of completion).

Next we explored the relationship between participants' overall gaze strategy and how much they benefited from social learning for lifting a heavy object. Figure 3.5E plots the preference score for the previously cued target object against the reduction in lifting duration in social conditions. Note that this measure of gaze preference in light weight trials reflects a baseline focus of attention as the demonstrator lifted the objects symmetrically in those trials and is only indirectly related to lifting scores which are obtained from heavy weight trials only. There is a slight correlation between those two measures in that participants who showed an increased preference for the cued object are somewhat more likely to exhibit increased social learning. However, this relationship can account for only a small fraction of the extensive variability in the dataset and is not significantly different from chance fluctuations (linear regression: $R^2 = 0.2$, $p = .11$) In particular, some participants (P6 and P7) showed pronounced social learning but only a small overall gaze preference as discussed above (figure 3.4).

Finally we compared the lift durations for heavy weight trials in social conditions for those trials when the heavy object was previously cued (7 trials per condition) against the few “heavy-switch” trials when the target object which had not been cued for 10±2 trials was lifted (1 trial per condition, of 4 switch trials). Participants lifting performance was not impaired in these heavy lift trials, and, unexpectedly, overall participants even showed a slight reduction in lifting duration when lifting a heavy target object in cue switch trials (paired t-test, mean duration 429ms for previously cued vs 388ms for uncued, $p = 0.004$).

As a reference, we show the lift-off durations for the demonstrator in figure 3.6. The lifting performance of the demonstrator was highly consistent for light (mean 116ms, std 60ms) and heavy lifts (mean 439.5ms, std 77ms), and also stayed consistent between experiments 1 and 2 (figure 3.6). Thus the robust delay for lifting heavy objects of more than 300ms provided a reliable cue to observers.
In debriefing, 10 participants reported that they thought watching the demonstrator lift first helped them in lifting heavy target object more smoothly. The remaining 4 participants (P3, P5, P6, P13) did not clearly identify such a relationship showed a weak (P3 and P13) to average (P5 and P6) social learning benefit.

Figure 3.6 Lift-off durations for all light lifts (n = 4704) and all heavy lifts (n = 896) performed by the demonstrator in experiment 1, as well as all light lifts (n = 1848) and all heavy lifts (n = 352) performed by the demonstrator in the social conditions in experiment 2.

3.4.2 Experiment 2

Participants' gaze behavior again showed high variability in their gaze behavior, both in social conditions as well as in passive conditions (see figure 3.7A-B). Unlike in experiment 1, participants' gaze behavior when observing the demonstrator lifting both target objects was symmetrical overall, again for both social as well as passive conditions. Participants did not show a preference towards the object of variable weight, neither in social (A, C) nor in passive conditions (B) even though we had expected this object to be more relevant for guiding their own actions.
Figure 3.7 Gaze behavior and lifting performance from all 11 participants in experiment 2. **A** Probability of gaze (mean over all light trials) being directed at the target object of variable weight (1), or at the target object of constant weight (-1) in light weight trials by participant (color-coded) in social conditions and **B** in passive conditions. (continued on next page)
Participants did show a significant benefit to their lifting performance, but the effect size was markedly reduced compared to Experiment 1. To investigate whether being able to predict the side that is going to be cued facilitated social learning of object weight as compared to the randomized cue schedule in experiment 2, we compared the size of the social learning benefit between experiments. A two-sample t-test revealed a significant contribution of cue predictability (group means of relative lift-off duration reduction in experiments 1 and 2: 167ms and 69ms, p = 0.0034).

In debriefing, only 3 participants (P1, P2, and P9) noticed that the weight was changing exclusively on one side for a given condition. To make sure that participants' lack of preference for the variable target object side was not merely the result of their not recognizing that weight changes were restricted to a given target object, we ran an additional group of participants on experiment 2 who were informed in advance about which object was going to be variable for each condition, again without showing a gaze preference for the variable object (data not shown).

3.5 Discussion

The introduction of competition for gaze during the demonstrator's bimanual lift of the target object resulted in high variability, with highly different gaze strategies being adopted by individual participants. Participants differed in the number and targets of gaze shifts in action observation, ranging from fixating mainly one target object or keeping their gaze on the central start object to shifting gaze back and forth between the two target objects multiple times, as well as whether observers shifted their gaze towards a heavy target object in response to a delay in lift-off. Overall there was a significant preference for directing gaze at the previously cued object in experiment 1, with participants again being highly variable in the extent of their preference, indicating a substantial effect of observer's own expected upcoming...
action on gaze allocation. We did not find any preference towards the object with variable weight schedule in experiment 2. Participants showed robust social learning benefits in their lifting performance which were stronger in experiment 1 than in experiment 2.

However, the stronger lifting performance in experiment 1 is not a straightforward reflection of participants knowing in advance which object they are going to lift themselves and selectively learning the weight of this object. In this case lifting performance would be diminished in cue switch trials, when participants expectations about which object to lift were violated. Unexpectedly, however, lifting performance was actually slightly enhanced in cue switch trials, perhaps as a consequence of participants' elevated attention after being surprised by the unexpected auditory cue and having to re-orient their attention towards the other object. Still, this result clearly shows that participants, even those who did not monitor the previously uncued object side with central vision, were at least as good at identifying a heavy weight at the side they generally attended less. This suggests that observers may have perceived the three weight configuration shown in the bimanual lift (both objects light, left object heavy, right object heavy) as a unit by picking up on asymmetries in the timing of lift-off between the two objects using peripheral vision.

This would explain the lack of preference for gaze towards the side of the object with a variable weight schedule in experiment 2, as such a preference would likely not have resulted in participants becoming more adept at lifting the occasionally heavy object. The experimental manipulation itself – arbitrarily setting one of two visually identical objects to a variable weight – may have been too artificial to be readily recognized by participants. Indeed, most of them did not even notice that the weight was changing on one side only in the first place. Therefore experiment 2 proved inconclusive as we did not succeed in making monitoring of the variable object relevant to the observer. The overall decreased benefit of social learning to participants' lifting performance in experiment 2 compared to experiment 1 may have been a consequence of reduced task engagement as participants were waiting for their cue telling them which object to lift.
Our results in experiment 1 clearly show a strong influence of action relevance on participants’ gaze behavior as participants overall strongly preferred to shift their gaze towards the object that was cued in the previous trials and was likely going to be their next target object for their own lifting action. These results generally support our proposal that the interests of observers in general, and their action goals in particular, are of key importance in action observation. Therefore experimental designs in which observers are asked to watch (video clips of) actors engaged in object manipulation passively for extended periods of time without themselves becoming active in turn clearly do not to capture the full range of gaze behavior in action observation (see also Dicks, Button, & Davids, 2010; Foulsham, Walker, & Kingstone, 2011). To address the effect of active engagement on the part of the observer we included a passive condition in our design but could do so only for experiment 2 as the logic of experiment 1 is based on repeatedly cueing a particular object for active lifting by the participant. Therefore the results in the passive condition remain inconclusive in the absence of a main result in experiment 2.

We have demonstrated that participants learn object properties such as weight from observing (Reichelt et al., 2013) and have proposed that this learning benefit may be a key factor accounting for why observers robustly shift their gaze to object contact events ahead of time (Flanagan et al., 2013). Thus we had expected participants to not only preferentially attend toward the object most relevant to their own subsequent action but also for their gaze behavior to translate into enhanced learning on this side. The latter was not the case. Instead, our results showed that the link between gaze position and social learning of object weight is more complex than we had initially assumed, at least for bimanual lifting of objects with unexpected weight. Without a direct connection between monitoring object contact events with central vision and subsequent lifting performance for that target object, it was not possible to evaluate the performance of different gaze strategies, all of which seemed to work well for participants overall. This may explain the striking variability of gaze behavior between participants in our task, as participants appeared to largely maintain whichever gaze strategies they had initially adopted over the course of the experiment.
3.6 References


Chapter 4

Predicting Choice Behavior in Action Observation

4.1 Abstract

This study examined the gaze behavior of observers who watched an actor's choice behavior in a virtual reaching task. The actor moved a cursor with a joystick first to a fixed start target and then to one of two choice targets, the locations of which were randomly varied across trials. In the 9:1 condition, one target was worth 9 points and the other was worth 1 point, and the actor always selected the high value target. In the 5:5 condition, both targets were worth 5 points, and the actor selected the closest target. We found that all participants shifted their gaze to the selected target before the cursor arrived. However, whereas some participants shifted their gaze to the selected target proactively (i.e., before they could use visual feedback of the cursor to determine the selected target), others exhibited a mixture of proactively and reactively initiated gaze shifts. Importantly, in almost all participants, the majority of proactive gaze shifts were directed to the selected target, and in those cases where observers shifted their gaze to an incorrect target first, the incorrect target was typically nearer to the start target and located close to the correct choice target. In a baseline condition where observers could not predict the correct choice target ahead of time observers either fixated the start target or, especially when the choice targets were close together, quickly shifted their gaze to the near target, with both strategies enabling them monitor the cursor motion and subsequently direct their gaze to the selected target often before the cursor arrived. Overall participants showed a mix of gaze strategies, based on prior knowledge, strategic guessing, and extrapolation, with high variability between and within participants. Taken together, these results demonstrate that observers naturally monitor the progress of the task with central vision, in line with previous work. The results also indicate that observers have the capacity, and often the inclination, to
exploit knowledge of the observed task, and the actor’s performance of the task, to predict the actor’s choice behavior in real-time.

4.2 Introduction

When watching other people perform familiar actions, observers – like actors – typically proactively shift their gaze to the forthcoming target location in time with movement onset to monitor object contact events and task progression (J. R Flanagan & Johansson, 2003; Flanagan, Rotman, Reichelt, & Johansson, 2013). However, in natural tasks, people often encounter situations where they select a particular object from multiple alternatives. Such decisions are governed by the values of the targets as well as the motor and temporal costs associated with attaining them (Cisek & Kalaska, 2010; Gold & Shadlen, 2007). Here we investigate how understanding the circumstances of an action translates into observers’ gaze behavior during action observation.

Most studies investigating gaze behavior in action observation either focus on fully predictable actions (J. R Flanagan & Johansson, 2003; Falck-Ytter, Gredeback, & von Hofsten, 2006; Rosander & von Hofsten, 2011) or actions whose target locations are not predictable from the outset (e.g. Ambrosini, Costantini, & Sinigaglia, 2011; Kanakogi & Itakura, 2011; Flanagan et al., 2013; but see Henrichs, Elsner, Elsner, Wilkinson, & Gredebaëck, 2014). Interestingly, when watching actors reach to unpredictable targets, observers still attempt to direct their gaze to the upcoming target in time for object contact. In a previous study (Rotman, Troje, Johansson, & Flanagan, 2006) investigating the influence of target predictability on gaze behavior: observers watched an actor first lifting a start block and then lifting either one of two target blocks positioned side by side. Participants exploited their knowledge of the actor's choice and shifted their gaze pro-actively to the target when the actor's choice was predictable, and switched to a reactive gaze strategy when the actor's choice was unpredictable. The study also showed an effect of target layout. When the target blocks were instead positioned along a line, participants in about half of the trials shifted their gaze strategically to the near target first and performed a second saccade to the far target when necessary.
Here we test the hypothesis that observers watching an actor selecting a target among alternatives will proactively position their gaze to upcoming object interaction events similar to the actor, based on the observers' understanding of the task rules and values adjusted to the demands of the current target configuration. We also investigate to what extent observers instead rely on reactively initiated gaze behavior exploiting movement direction cues, and on strategic gaze behavior exploiting the spatial target configuration when the actor's target preference is ambiguous. In our task the actor used a joystick to move a cursor on a screen from a start zone to an end zone via a fixed start target and either one of two choice targets whose positions were varied from trial to trial. The values associated with the choice targets were varied across conditions: in the 5:5 conditions, both targets were equally valuable elevating the importance of the spatial configuration for target choice. In the 9:1 condition, one choice target was considerably more valuable than the other. We also ran a baseline condition in which the actor's choice was completely random as well as a single target condition as reference. We recorded the eye movements of observers, who were informed about the values of the choice targets at the beginning of each condition.

4.3 Methods

4.3.1 Participants

One actor (man, 29 years) and nine observers (5 women, 18-27 years) participated in this study. All participants gave informed consent and were uninformed about the research question. The local university ethics board approved the study, which complied with the Declaration of Helsinki.

4.3.2 Apparatus

The actor was seated in front of a table and moved a cursor on a vertical screen by applying forces to a spherical joystick held in the preferred hand (see figure 1A). The joystick was mounted on a six-axis force/torque transducer (FT-Nano 17; Assurance Technologies, Garner, NC) that measured the forces in the horizontal plane at 500 samples/s. Targets and cursor position were projected onto a screen (1.1 by 1.35 m) located 1.75 m in front of the actor’s eyes. In observer trials, the participant was seated in the
same position (see figure 4.1B) and the joystick was moved over to the experimenter sitting to the left of the observer. The task, shown in figure 1C, consisted of moving the cursor from the start to the end zone (both 6 by 6 cm) via the start target (radius of 3 cm) and one of the two choice targets (radius of 2.5 cm). The vertical distance between the start and end zones was 0.82 m, corresponding to 26° of visual angle. For each trial, the two choice targets were randomly positioned at least 15.8 cm apart on a horizontal line spanning 57 cm (16.7° of visual angle) to either side of the midline and located 9.7 cm above the center of the screen. The position of the cursor scaled linearly with force applied in the horizontal plane, with 1 N of force moving the cursor 5.3 cm. The force signals controlling cursor were low-pass filtered at 3 Hz to prevent cursor wobble driven by the actor’s physiological tremor (Säfström, Flanagan, & Johansson, 2013). The actor's and later the observers' gaze positions were recorded using an infrared video-based eye-tracking system (RK-630PCI pupil/corneal tracking system; ISCAN, Inc., Burlington, MA) in the vertical plane at 240 samples/s. The eye tracker was mounted on a steel frame fixed to the table. Participants leaned their foreheads against a leather band and placed their heads on a chin rest to keep orientation stable. We used a streamlined version of the calibration sequence described in (Johansson, Westling, Bäckström, & Flanagan, 2001) using a direct linear mapping of the differential between the pupil and corneal reflection position signals to screen coordinates based on a 9 point calibration sequence.
**Figure 4.1** A Top view of actor trials. B Top view of observer trials. C Task layout in an example trial showing start and end zones and start target (green), and choice targets (here in blue and red) with the cursor trajectory overlaid. Choice targets were positioned at variable locations within the choice target zone in every trial. The size of Cursor and Gaze regions, only used for analysis, are shown in dashed lines.

### 4.3.3 Procedure

**Actor trials.** At the start of each trial, the start zone appeared at the bottom of the screen. After the actor held the cursor within the start zone for 250 ms, the targets and the end zone were presented and the actor immediately moved the cursor first to the start target, then to one of the choice targets, and finally to
the end zone. The actor needed to hold the cursor within a target for 150 ms to successfully attain it which was signaled by an auditory beep. The score of a trial was computed as the value of the selected choice target divided by total trial time and was shown to the actor together with the running average of the current condition at the end of each trial. The cursor trajectories of the actor were recorded to be played back to participants in observer trials.

**Observer trials.** Participants first experienced moving the cursor on an empty screen themselves. Then the joystick was moved over to the experimenter and the participants were instructed to watch the experimenter performing the task on the screen. The experimenter first gave a live demonstration of 3 example trials after which the eye tracker was calibrated and the cursor control was switched to previously recorded playback trials while the experimenter only pretended to perform the task live. The experimenter kept his hand on the joystick and observers who could see the experimenter only out of the corner of their eye could not tell that we used video playback due to the isometric force-to-position mapping. We pretended to run the trials live so that participants who were passive observers throughout would feel more engaged while at the same time being able to present identical trial videos to participants rather than introducing variability on the part of the actor. We omitted feedback scores in all trials shown to participants.

**4.3.4 Conditions**

The actor was instructed to maximize the score in every trial and was informed of the set of target values in each condition. The actor received extensive training to ensure consistent behavior within and between conditions before recording. Observers were likewise told the current set of values at the beginning of each condition.

**5:5 condition:** Both targets were of equal value (5 points each) and colored purple. We expected the actor to consistently choose to reach for the choice target closest to the midline in order to minimize cursor trajectory and movement time and to choose a target at random in equidistant target configurations.
9:1 condition: Target 1 held high value (9 points) whereas target 2 was only worth 1 point. For the actor, the high value target was consistently colored red and the low target was colored blue, while for half of the observers the choice target colors were switched in playback to control for potential preferences for color. We expected the actor to choose the high value target exclusively in order to achieve high trial scores.

Baseline condition: This condition was constructed manually based on trials from a separate recording of a 9:1 condition performed by the actor. In replay, the color of both choice targets was set to purple, resulting in effectively random choice behavior for observers who were told that the actor's choice behavior would be random in this condition.

Single target condition: Here only only one purple target was presented worth 5 points which the actor had to attain to complete the trial.

Observers were first shown the 5:5 and 9:1 conditions in counterbalanced order, and were then shown the baseline and single target conditions also in counterbalanced order (see Table 4.1 for a graphical illustration). We did not completely randomize the order of conditions to ensure observers would approach the 5:5 and 9:1 conditions intuitively before encountering arbitrary choice behavior in the baseline condition.

Table 4.1: Graphical illustration of the order of experimental conditions (left). Double-arrows indicate that conditions were presented in counterbalanced order between participants. As a reminder we also provide a list of the logic behind the actor choice behavior for each condition (right). Details see text.
4.3.5 Data Analysis

Cursor position and gaze signals were sampled at 500 Hz and 240 Hz, respectively, and smoothed using a fourth-order, zero-phase lag, low-pass Butterworth filter with a cutoff frequency of 14 Hz and 25 Hz, respectively. For analysis, we used the cursor position computed in real time from the force signals, as well as gaze positions synchronized to the cursor trajectory in the respective trial. All signals were interpolated to 1000 Hz. Regions of interest around the target positions were used to compute timing of events in a trial. Target regions for cursor arrival and exit had a radius of 5.3 cm and gaze area of interests had a radius of 10.6 cm corresponding to 3.5° of visual angle (see figure 4.1C).

Trials were categorized as proactive when observers shifted their gaze to any one of the choice targets within 150 ms after cursor exit from the start target (see results: observer gaze behavior, for a derivation of the 150 ms criterion) and as reactive otherwise. Trials were categorized as correct when the first choice targets looked at by the observer was the actual target attained by the actor for that trial, and otherwise as incorrect. Likewise, trials were categorized as near trials when gaze was first shifted to the target closer to the midline. In a few trials observers looked at both choice targets ahead of the cursor leaving the start target; in this case, only the choice targets looked at after cursor exit from the start target were considered for analysis.

To measure how well individual observers were able to correctly predict the upcoming target choice, we computed the proportion of correct trials, as well as the proportion of near trials, for all proactive trials and conducted binomial tests for each observer and condition. To assess the effects of conditions at the group level, we used paired t-tests. The Holm-Bonferroni test was used to correct for multiple comparisons. A p-value of less than .05 was considered statistically significant in both tests.
4.4 Results

4.4.1 Actor performance and choice behavior

Figure 4.2B shows the latencies for when the cursor and the actor's gaze exited the start target zone and arrived at the choice target zone relative to when the start target was attained – i.e. when the cursor was successfully held within the start target for 150 ms – as cumulative frequency distributions, expressed at proportions, for all of the trials performed by the actor. On average, gaze exited the start target about 150 ms after target attainment and the cursor exited about 30 ms later. The results indicate that the actor generally anticipated the time at which the start target would be attained and did not wait for auditory feedback signaling attainment of the start target before initiating his eye and cursor movement to the choice target, in which case we would expect a delay of about 300 ms between gaze and cursor exit from the start target and start target attainment (e.g. Säfström et al., 2013). On average, gaze entered the choice target zone around 90 ms after exit from the start target, consistent with a single saccadic eye movement between the targets. Movement duration of the cursor to the choice target was more variable, reflecting variable target placement, and lasted on average about 300 ms.
Figure 4.2 A-C Scatterplots of target configuration in individual trials shown as the distance from midline of choice target 1 minus the distance from midline of choice target 2 as a function of trial number, for the 5:5 condition (A), for the 9:1 condition (B), as well as for the baseline condition (C). Each trial is color-coded to indicate which target the actor selected. D Cumulative proportions of the timings of actor's cursor and gaze departure from the start target and arrival at the selected choice target, aligned to the time at which the start target was attained. Data from all trials performed by the actor are included.

Figure 4.2A-C shows the actor's target selection in the 5:5, 9:1, and baseline condition for all 70 trials as a function of target placement expressed as the difference between the distance of target 1 and target 2 from the midline. In the 5:5 condition the actor selected the target that was closer to the midline (target 1 for positive and target 2 for negative values for distance to midline T2-T1) in all but 3 trials in which the target distances were similar. Thus, in the 5:5 condition, the actor’s choice was generally predictable for observers except in a subset of ambiguous trials where choice target locations were almost equidistant. In
the 9:1 condition, the actor always selected the more valuable target (T1), indicated by target color, regardless of which target was closer to midline. (Note that in a control experiment not described here we found very similar behaviour across different actors.). In the baseline condition, observers were shown arbitrary, and hence unpredictable, choice behavior.

4.4.2 Observer gaze behavior

4.4.2.1 Deriving the criterion for proactive gaze behavior empirically

To differentiate proactive from reactive eye movements we first determined the point in time when observers can effectively integrate sensory information to generate eye movements to the correct target reactively. Figures 4.3A-B show gaze arrival times at the first fixated choice target, relative to the cursor exit from the start target, in the baseline condition where no prior information about the actor's choice was available for an example participant (P6, A) and all participants (B). Next we computed observers' gaze arrival times at the correct target – the target chosen by the actor – and contrast it with the cumulative proportion of the first target halved which represents the proportion of correct target choices which would be expected by chance given there were 2 possible targets (see figure 4.3C which shows these distributions for participant P6). The difference between these cumulative proportions, shown in figure 4.3C for participant P6, is a measure of the observer's cumulative target estimates with respect to chance. A value of 0 represents the observer looking at the correct target at chance level, the maximum value of 0.5 represents looking at the correct target in all 70 trials. To characterize gaze behavior functionally with respect to effective monitoring of the choice target, this measure also includes incorrect trials, where observers first shifted their gaze to the incorrect target and then performed an additional, corrective eye movement to the correct target. As shown in figure 4.3E, participants' estimates only begin to exceed random choice consistently at about 200 ms after cursor exit from the start target reflecting information about cursor movement trajectory becoming available. To preclude the possibility of online information about the current movement being used to guide the decision we chose a conservative criterion and defined all trials where participants shifted their gaze to any one of the choice targets before 150 ms after
cursor exit (see vertical dashed gray line in Fig. 4.3E) as *proactive* and, conversely, all trials where their eye movements arrive later as (potentially) *reactive*. Note that this measure is based on gaze relative to action onset to distinguish between eye movements initiated based on anticipation of the upcoming target location as opposed to extrapolation of an ongoing movement trajectory. In contrast, predictive gaze is usually defined relative to action completion (i.e. cursor or hand arriving at a given target) in the literature (e.g. Falck-Ytter et al., 2006; Henrichs et al., 2014).
Figure 4.3 Cumulative Proportions of gaze arrival times over all trials in the baseline condition for an example participant (P6, left: A, C, D) and for all 9 participants (colored traces, right: B, E). The panels show gaze arrival at any of the choice targets (*First Target*, A, B), at the correct target (C), and the correct target above chance computed as the difference between the cumulative proportion between gaze arrivals at the correct target and gaze arrivals at the first target halved (D, E). This measure is illustrated for participant P6 by the shaded areas (C, D). The gray dashed line bottom right panel denotes the 150 ms criterion used to distinguish proactive and reactive trials.
4.4.2.2 Individual results

Figure 4.4 shows, for each condition (rows) and each participant (colors), the cumulative distributions of gaze arrival at any choice target (left column), and the proportion of proactive trials (trials where gaze is directed at a target before 150 ms after cursor exit (vertical dashed gray line in first column). There was considerable variability in gaze latencies to the first choice target between conditions as well as between participants in a given condition: some participants tended to generate proactive eye movements in most trials in all conditions (P2, P6, P8, and P9), with two of them often looking at one of the choice targets well ahead of time, before the actor even attained the start target (P2, and P9). Two participants generated considerably more proactive saccades in the baseline condition (P1 and P4) whereas still other participants were stable over conditions (e.g. P6 and P8). Given this variability, and since participants generated a reasonable number of proactive trials in all conditions, we opted to first assess each participant separately. Specifically, we used binomial tests to determine whether the proportion of proactive trials in which gaze was first directed to the correct target, was significantly different from chance (Correct | Proactive, third column, filled bars indicate $p < 0.05$). We also determined the proportion of proactive trials where observers first looked at the target closer to the midline (Near | Proactive, right column, filled bars again indicate $p < 0.05$).
In the 5:5 condition (first row in figure 4.4), participants typically showed proactive gaze behavior in the majority of trials as expected. About half of the participants (P1, P2, P6, P8, and P9) were proactive more than two-thirds of the time while others (P3, P4, P5, and P7) were proactive in about half of the trials. All participants showed a significant preference for the near target which was almost always the correct target in this condition with around half of the participants being correct in virtually all trials and all participants being correct in at least two-thirds of the trials in which they shifted their gaze proactively.

In the 9:1 condition, participants also showed proactive gaze behavior in a clear majority of trials. Some (P1, P2, and P8) did so in almost all trials whereas others (P3, P4, P5 and P7) were proactive in only about half of the trials with the remaining participants (P6, P9) in between. All participants tended to look at the correct target in proactive trials, ranging from 60 percent to nearly all trials, with 7 of 9 participants looking at the correct target significantly above chance. All participants showed at least a slight preference for the near target in proactive trials, significant in 4 participants, even though in this condition the actor showed no such preference in his choice behavior and always selected the more valuable target irrespective of target positions.

In the baseline condition where the actor's choice behavior was unpredictable, all participants still shifted their gaze proactively to one of the targets in a subset of trials and 5 of them (P2, P3, P6, P8, and P9) did so in a majority of trials. While a few participants showed reactive gaze behavior much more frequently in this condition as compared to the 5:5 and 9:1 conditions (e.g. P1, as well as P4 and P5) most
participants still generated about the same proportion of proactive trials. When proactive, participants were guessing at chance levels and all 9 participants directed their gaze significantly more often to the target closer to the midline.

In the single target condition, participants predominantly adopted a proactive gaze strategy as expected and typically shifted their gaze to the single available target either in time with or slightly after cursor exit in a majority of trials with some (P1 and P8) doing so in almost all trials and one participant (P4) in only about half the trials.

4.4.2.3 Group results

Figure 4.5 shows the mean proportions of proactive trials for all 9 participants for all conditions, and of those the proportions correct and near, respectively. To compare conditions we carried out paired t-tests using the Holm-Bonferroni correction to adjust \( p \) values. On average participants were proactive in two-thirds of the trials in both the 5:5 and 9:1 conditions and were somewhat less likely to be proactive in the baseline and more likely to be proactive in the single target condition. However, no significant differences between pairs of conditions were observed in the proportion of proactive trials (\( p > 0.12 \) in all 6 cases). The proportion of correct | proactive trials was significantly greater in the 5:5 condition than in the baseline condition, \( t(8) = 7.50, p = .003 \), and significantly greater in the 9:1 condition than in the baseline condition, \( t(8) = 6.30, p = .002 \). There was no difference between the 5:5 and 9:1 conditions, \( t(8) = 1.07, p = .318 \). Participants looked to the near target in about 8 out of 10 trials on average in proactive trials in the 5:5 and the baseline condition. In the 9:1 condition they did so significantly less frequently as compared to the 5:5 (\( t(8) = 3.64 \) and \( p = .014 \)) and the baseline conditions (\( t(8) = 4.65 \) and \( p = 0.006 \)). Overall, participants were on average proactive and correct in around 60 percent of trials in the 5:5 and 9:1 conditions and, as expected, in slightly less than 30 percent of trials in the baseline condition.
Next we asked how the distributions of observer “errors” – that is, trials in which participants first looked to the incorrect object in the 9:1 and 5:5 conditions – relates to target configurations. Figure 4.6 shows the proportion of trials in which participants shifted their gaze to target 1 as a function of the relative distance from the midline of target 1 and target 2 and of configuration type (dark gray when both targets were placed on the same side of the midline, light gray when placed on opposite sides). As expected, participants accurately chose the near target in the 5:5 condition for configurations where one target was clearly closer to the midline in almost all trials, and became less discriminant for increasingly ambiguous trials (relative difference close to 0). A similar distribution was found in the baseline condition. In the 9:1 condition, participants almost always shifted their gaze correctly to the higher value target first when this target was positioned closer or equidistant to the midline. Gaze behavior in
configurations where the high value target was located further away than its counterpart was more variable. Crucially, in such configurations, participants are much more likely to look at the incorrect targets first when (1) both targets were on the same side and (2) the incorrect target is nearer. Even in reactive 9:1 trials, almost all participants clearly prefer to look at the incorrect target first for the configurations where the incorrect target was located adjacent to the correct target but closer to the midline and thus nearer to the start target. In contrast, in the 5:5 condition, gaze is almost always directed at the correct near target especially in reactive trials, whereas in the baseline condition, proactive gaze (guessing) is directed at the near target and reactive gaze shows no spatial pattern, reflecting gaze shifts to the correct target (target choice being unrelated to configuration in this condition).
Figure 4.6 Scatterplots showing the mean proportion of trials where participants looked to target 1 first binned by target configuration expressed as the difference between the absolute distances from the midline of both targets for the 5:5, 9:1, and baseline conditions (rows) for proactive and reactive trials (columns). Dark gray circles indicate configurations where both targets were located on the same side whereas light gray circles indicate configurations where the 2 targets were located on opposite sides of the midline.

4.4.2.4 Visual consequences of gaze shifts

Finally, we also determined when the participants' gaze arrived at the correct choice target relative to cursor arrival at this target (as opposed to cursor exit). This is shown in figure 4.7 expressed as cumulative proportions for all conditions. Participants' gaze arrived at the correct target before or in time with the cursor in nearly all trials in the single target condition, in around 80 percent of trials in both the 5:5 and 9:1 conditions, and still in over 60 percent of trials in the baseline condition (brown traces). This effectively enables monitoring of target contact in central vision in most trials in all conditions, and of target attainment in almost all trials (attainment occurs at minimum 250 ms after target contact, recall that to attain the target, the cursor had to be held within the target for 250 ms, see dark gray traces). Note that it is quite possible for gaze in reactive trials to still overtake the cursor en route to the target and arrive there earlier. This holds true for more than half of reactive trials in the baseline condition, showing that fixating the start target and monitoring the outward cursor trajectory was effective for disambiguating target choice. However, proactively shifting gaze to one of the targets (typically the near one, see figure 4.6 above) appears to be at least equally effective. When directed at the correct target (blue traces), proactive gaze shifts obviously arrive very early allowing monitoring of target approach and contact. Even when directed at the incorrect target (pink traces), i.e. in on average every second trial in the baseline condition, the cursor trajectory to the other choice target can be monitored and a corrective eye movement initiated resulting in an only slight delay compared to reactively initiated trials, still allowing monitoring of target attainment if not target contact.
Figure 4.7 Cumulative proportion of gaze arrival times at the correct choice target relative to cursor arrival at the same target for all participants and trials in all conditions (rows). To compare the consequences of proactively vs. reactively initiated gaze shifts, proactive trials were split into correct trials (proactive +, green trace) and incorrect trials (proactive -, pink trace, reflecting an additional corrective eye movement). Reactive trials are shown in blue, and overall gaze arrival for all trials is shown in brown as reference. The first black dashed line denotes concurrent arrival of gaze and cursor (0 ms), the second dashed line indicated the earliest possible target attainment time (250 ms). The dark gray trace shows the actual distribution of target attainment times by the actor.
4.5 Discussion

Eye tracking studies of action observation reliably show that observers - like actors - attempt to shift their gaze to the target location ahead of time to monitor target contact and attainment (Falck-Ytter et al., 2006; Flanagan & Johansson, 2003; Henrichs et al., 2014; Rosander & von Hofsten, 2011; Rotman et al., 2006). In this experiment, we varied choice target configurations systematically to investigate whether observers exploit their knowledge of the actor's task set (values and movement costs for attaining the choice targets) to predict the actor's choice behavior. Our results clearly show that observers were capable of exploiting their prior knowledge about the circumstances of an action when available, and reliably showed proactive eye movements to the target that would be chosen by the actor. In the 5:5 condition, we expected participants to be aware of the actor choosing the near target with high likelihood given equal target scores, and all 9 participants appropriately shifted their gaze proactively to the near target in a majority of trials. However, participants also frequently generated proactive eye movements almost always directed at the near target in the baseline condition, with only a few participants using a reactive strategy, resulting in highly similar gaze behavior between these two conditions. Hence we cannot strictly speaking rule out the possibility that participants' performance in the 5:5 condition was (partly) based on guessing rather than exclusively guided by their prior knowledge. However, in the 9:1 condition, participants clearly deviated from the near-target, directing their gaze to the high value target ahead of time in a majority of trials overall clearly guided by their knowledge of the circumstances of the action. Taking a closer look, participants gaze behavior in this condition also incorporated strategic elements, in that when observers shifted their gaze to the low-value target first, this target was positioned en route to the high value target in the majority of those trials. Taken together, our results clearly show that observers have the capacity, and often the inclination, to exploit prior knowledge about the circumstances of an action to predict the actor’s choice behavior.

We had initially expected participants to more frequently rely on reactive gaze behavior in the baseline condition where target choice was unpredictable, as they did in a previous study (Rotman et al., 2006) for
choice targets positioned on the sides of a start target. In that study, however, target positions were not varied systematically to cover the full range of configurations we examined here. Proactive guessing in the baseline condition turned out to be a viable strategy in the baseline condition: In every second on average, it delivered hits on the correct target early on and in the other half of trials when the incorrect target was chosen, the subsequent corrective response incurred a temporal cost small enough to eventually arrive at the correct target at least in time to monitor target attainment. Thus participants combined a variety of gaze strategies, guided by prior knowledge of the circumstances of the action as well as strategically exploiting target configuration and utilizing on-line information about cursor movement trajectory, appropriate to the task conditions.

One of the most striking findings of this study was the presence of considerable variability between participants who differed both in terms of how frequently they would commit to a proactive gaze strategy and to the extent they directed their gaze to the correct target when doing so in the 5:5 and 9:1 conditions. In particular, participants adopted a different mix of gaze strategies: when the target was predictable, some participants typically shifted their gaze to the choice target proactively in a similar manner as the actor as originally reported (Flanagan & Johansson, 2003), while other participants showed a mix of proactive and reactive gaze strategies. Still others, particularly P2, P9, and sometimes P6, directed their gaze to a target even earlier, bypassing the start target altogether. To our knowledge this is the first study to systematically describe and relate individual variability to differential reliance on gaze strategies for action observation.

There are 3 main factors that may account for this variation. First, the design of our study introduced a spectrum of target predictability and target configurations. In spite of their variability in gaze strategies, all participants effectively managed to monitor target contact and attainment, arriving at the target location ahead of or in time with the cursor in a majority of trials even in the baseline condition. Clearly the metrics of the target choice task – i.e. the timing and duration of the actor's reaching movements – allowed for target prediction, strategic guessing, as well as reactive gaze strategies to be generally successful. While participants overall did adjust their gaze strategies to the demands of a given condition...
and were on target earlier on average in the 5:5 and 9:1 conditions compared to the baseline condition, our design allowed participants greater flexibility to adopt a gaze strategy of their choice compared to studies whose design favors or even allows for reactive gaze strategies only. Second, participants were instructed to simply watch the actor perform the target choice task and may have arrived at very different interpretations about what exactly they were asked to do. We have previously shown that participants may adopt radically different gaze strategies in response to particular task instructions (Flanagan et al., 2013). Finally, observers appeared to differ considerably in terms of their motivation to participate in the experiment. Watching video recordings is likely to be less engaging than watching a live actor (e.g. Dicks, Button, & Davids, 2010), and hitting virtual targets is likely less effective at drawing an observer's attention compared to real targets, which could have made it easier for participants to disengage their gaze from the cursor contacting the start target (compare Gesierich, Bruzzo, Ottoboni, & Finos, 2008 who reported high variability between observers when watching a virtual version of the block stacking task).

While highly simplified, observing an actor engaged in choosing one of two virtual targets does capture key aspects of the structure of real world action observation, as we regularly observe other people engage in familiar activities without always being able to predict their exact next move with certainty. Studies of social learning by observing an agent engaged in decision making under uncertainty typically present observers with only abstract, categorical choices and outcomes (see e.g. Burke, Tobler, Baddeley, & Schultz, 2010) so their designs do not include presenting the actual actions which constitute choice behavior to obtain rewards, even though actions and their temporal and motor costs are an important aspect of decision making and appreciated intuitively by actors (Cisek & Kalaska, 2010; Cos, Duque, & Cisek, 2014). While we here include the actor's movements as part of the decision process, we did not address social learning of object value as participants were explicitly informed about task structure and target values in advance. However, our results lay the groundwork for an investigation into how observers may learn about object values by observing choice behavior of others, by showing how prior knowledge about the circumstances of the action translates into observer gaze behavior. In particular, our results show that proactive gaze shifts should only be used as an indicator for prior knowledge with caution, as
gaze behavior will not always reflect prior knowledge directly, especially for spatial configurations which favor guessing strategies.

This study also illustrates the crucial importance of clearly differentiating between proactively initiated eye movements and reactively initiated eye movements guided by information about movement trajectory which still “proactively” arrive at the target before or in time with the actor. Unfortunately, action observation studies routinely report only (mean) gaze arrival at the target area of interest (e.g. Elsner, D’Ausilio, Gredebäck, Falek-Ytter, & Fadiga, 2013; Möller, Zimmer, & Aschersleben, 2014), which has become the “standard measure used in the majority of eye-tracking studies with infants” in particular (Henrichs et al., 2014; but see Rosander & von Hofsten, 2011). While gaze arrival at the target relative to the hand/cursor is certainly an important functional measure for understanding the effectiveness of observer's gaze behavior as discussed above, the practice of exclusively relying on this measure for classifying eye movements as “predictive” is deeply problematic as it glosses over many important aspects of the data (compare figure 4.4 to figure 4.7), and obscures different gaze strategies the observer may have been relying on. In particular, latencies for arriving at the target relative to the actor depend on the movement metrics of the actor and the spatial configuration of the task. As eye movements are considerably faster than reaching movements, the observer's gaze may easily overtake the actor's hand/cursor en route to the target based on an online estimate of target trajectory. The duration of the temporal window which still allows the observer to achieve this crucially depends on the actor's movement duration which may vary widely between studies (ranging from less than 0.5 s in J. R Flanagan & Johansson, 2003; to 1.5 s for Kanakogi & Itakura, 2011). Therefore it is crucial that researchers report the full range of movement metrics and the point in time when observers begin to effectively utilize movement information reactively in real time, in order to more fully appreciate the gaze strategies used, interpret effects, and clarify contradictory reports, in particular when information about the timing of eye movement initiation is not separately reported. As knowing about the circumstances of an action does not necessarily directly translate into proactive gaze behavior in action observation, conversely, proactive gaze initiation on the part of an observer, let alone proactive gaze arrival at a target, is not necessarily a
direct marker for understanding the structure of a task but requires a detailed characterization of the task metrics, its spatial configuration, and the effectiveness of different gaze strategies for further interpretation.

4.6 References


Flanagan, J. R., Rotman, G., Reichelt, A. F., & Johansson, R. S. (2013). The role of observers’ gaze behaviour when watching object manipulation tasks: predicting and evaluating the consequences of


Chapter 5

General Discussion

5.1 Chapter summaries

5.1.1 Social learning of object weight (chapters 2 and 3)

In our first study (chapter 2) we quantified social motor learning by observing for the case of object weight and showed that while observers substantially update their lift forces following action observation, they do not do so to quite the same extent compared to after they lifted the object themselves. This result is important for three related reasons: First, social motor learning of object properties has not been described quantitatively before, as research in motor learning has instead focused on social learning of novel dynamics in reaching (Mattar & Gribble, 2005; Malfait et al., 2010), and the social cognition literature typically views social learning as learning by insight on the level of action selection (e.g. Heyes & Galef 1996, Huber 2012). Second, intuitive learning of object weight from observing supports our view that extracting information about object properties is part of why observers, like actors, monitor contact events. Third, it demonstrates that object lifting can serve as a model system to study motor learning by observing and provides a baseline for studies with more complex designs.

The second study (chapter 3) introduces a design to study the effects of the observer's action goals on their gaze behavior in action observation and to describe the consequences of gaze behavior on social motor learning. We report a substantial effect of the observer's expected action goals on their gaze behavior, but did not find an enhanced social learning advantage when participants lifted the object which they had monitored preferentially before. No effect on gaze was present in the second experiment that focused on object properties (variably weight schedule). In the light of the results from experiment 1, which suggest that there is no clear-cut relationship between monitoring a given object with central vision and learning about object weight by observing simultaneous bimanual lifting of two objects, this is
perhaps not surprising. Moreover, variability of weight as an object property may not be intuitively picked up on by observers, rendering the negative result in experiment 2 inconclusive rather than a refutation of the proposal that action goals of observers drive their gaze behavior in action observation.

We did not fully appreciate, a priori, the complexities involved in observing and learning from bimanual lifting of two objects simultaneously. The high variability between participants introduced by simultaneous lifting of two objects raises the question of the benefits of central and peripheral vision for extracting information about object properties for a given action. This points to the possibility that there may be a range of potential gaze strategies for observers for more complex activities. We have previously demonstrated that explicit instructions to observers can fundamentally change their gaze behavior when observing a simple object lifting task (Flanagan et al. 2013), and our present study at least illustrates the challenges in trying to manipulate the intrinsic motivation of observers in action observation as well as in evaluating the effects of gaze behavior in settings which involve competition for gaze.

5.1.2 Social learning of object value (chapter 4)

Object value is a relational concept that is informed by the intrinsic properties of a particular object but also reflects goals, needs, and inclinations of a person in a given situation. The study presented in chapter 4 lays the groundwork for investigating social learning of object value by describing in detail how knowledge of the circumstances of an action translates into gaze behavior in action observation. Switching to a virtual design allowed us to present the two potential choice targets in a range of spatial configurations but may have also resulted in a higher degree of variability of gaze behavior both within and between participants.

Even though we kept the design of the study as simple as possible – first by telling observers about the values explicitly beforehand, and second by only including conditions where either the value of the object (9:1 condition) or the spatial target configuration (5:5 condition) exclusively determined the actor's choice behavior – observer gaze behavior showed remarkable complexity and individual differences which we interpret in terms of a mixture of gaze strategies (anticipating, guessing, reacting). Unlike the studies
presented in the previous chapters – but similar to Rotman et al. (2006) and indeed most of the studies on action observation in the literature – participants in this study were only passively observing the cursor hitting the virtual objects throughout the task in the absence of a particular task (apart from looking at the screen). From a motor control perspective, the observation that observers generate qualitatively similar gaze behavior in such conditions compared to when they intermittently interact with objects themselves rather than e.g. fixating the middle of the display, itself underscores that to observers find keeping track of other people's actions in the environment to be rewarding on its own terms.

Our results clearly show that target configuration has a systematic effect on gaze behavior well beyond the differences demonstrated by Rotman et al. (2006) between two extreme cases (illustrated in figure 1.7). We show that prior knowledge of the action goal largely determines the mode of gaze behavior. However, target configuration, which is the main factor when the actor's choice is unpredictable, can modulate and even override the effects of prior knowledge. Finally, the choice of gaze strategy between anticipation and extrapolation is likely modulated by task engagement, in particular in designs based on observation of virtual objects or video clips which appear to be less engaging for observers compared to real object manipulation. Taken together, the results of this study were instrumental for developing the conceptual framework introduced in the introduction and applied in the following sections of the general conclusions. Moreover, we introduce empirical quantification of the effective latency criterion for differentiating between predictive gaze behavior based on anticipation and extrapolation. We also present simple analysis tools for describing gaze strategies on the single trial level, by participant, and on the group level, identifying the mode of prediction defined relative to action onset and evaluating visual consequences at action arrival. Thus we show how the exclusive focus on main effects in summary statistics can be misleading – which unfortunately is the practice adopted by many studies in the field, as discussed below.
5.2 Applying the modes of predictive gaze and the framework of contextual factors to critically discuss the predictive gaze literature

5.2.1 Overview

This section begins with a concise treatment of the concrete implications of the framework of contextual factors shaping predictive gaze behavior, to develop a general intuition of what kind of gaze strategies to expect in a given situation or experimental design. This integrated perspective then provides the basis for a systematic review of the current literature on predictive gaze in action observation. Up to now, a critical discussion or even overview of the field has been missing, instead the practice has been to dutifully summarize the results of studies in generic terms. Therefore, I felt it necessary to engage in much more precise analysis of research in the field, and also to draw more general methodological lessons so as to not get completely bogged down in details.

The section is organized in 3 parts. I start out with a review of selected studies of predictive gaze behavior which address contextual factors, revealing a clear pattern: while the original report by Flanagan & Johansson (2003) is frequently cited as “seminal”, most studies on predictive gaze in action observation are instead based on a methodology which treats gaze behavior (now defined in relation to gaze arrival rather than initiation) akin to reaction times. The scope and limitations of this methodology are explicitly spelled out.

Next I revisit the relationship between predictive gaze behavior and the mirror neuron system, which while constantly appealed to in the literature had remained unclear until very recent studies have begun to shed some light on this question. This section is organized around the distinction between anticipation based on prior knowledge and prediction based on extrapolation of a currently visible action. The distinction between these different modes of prediction has played an important role in the debates about mirror neuron system function quite independent of its relation to predictive gaze behavior.

The final section is dedicated to a systematic discussion and literature review of studies addressing the role of familiarity and expertise in action observation, in particular the proposal that a basic motor ability
to perform a given action is necessary for engaging in predictive gaze behavior when observing this action resulting in a close correspondence between the developmental trajectories of action and action observation.

This thesis itself is written primarily from the perspective of motor control, and is based on research into eye-hand coordination and object manipulation, extending and organizing the results of studies from our lab beginning with the original demonstration by Flanagan & Johansson (2003). Most of the other studies reviewed here draw much more heavily from cognitive (developmental) psychology, where researchers are used to looking at behavior either as the consequence or as an indicator of cognitive processes. Their methodology and especially their style of reporting results in places are quite different from those within the field of motor control. In particular when modes of predictive gaze are often conflated, making it very difficult to reconstruct the actual gaze behavior that observers engaged in. This problem is in my opinion the consequence of a lack of engagement and communication between research traditions. A more balanced discussion of what motor control, cognitive psychology, as well as other fields and research programs addressing social interaction each have to contribute, and how to relate and eventually integrate their concerns and concepts is sorely needed. However, apart from a brief discussion at the end of this chapter, this exceeds the scope of the present work. The main contribution of this chapter is to apply the framework developed in the introduction to provide concrete suggestions about what gaze behavior to expect and consequently how to design, report, and interpret studies, as well as how to integrate insights into predictive gaze behavior in action observation.

5.2.2 Modes of predictive gaze in the literature

Several authors at times effectively distinguish between what is referred to here as distinct modes of predictive gaze behavior in action observation (anticipation, extrapolation, and tracking, see introduction section 1.4.1). For example, Ambrosini and colleagues (2011) have deliberately set up their design to facilitate eye movements based on extrapolation rather than anticipation. Rosander and von Hofsten (2011) explicitly set out to differentiate between proactive and reactive gaze initiation as well as arrival,
but ended up reporting only gaze behavior based on extrapolation in all their conditions (presumably due to the variable timing of movement onset, see section 5.2.3.1 below). Indeed, most recent studies on “predictive gaze” are focused on gaze behavior based on extrapolation (it is hard to be certain as typically only mean gaze latency at arrival is reported). This conflation about modes of prediction is most clearly shown when the “original” or “seminal” study (Flanagan & Johansson, 2003) is cited in support of definitions of prediction in terms of extrapolation either implicitly (Elsner, D’Ausilio, Gredebäck, Falck-Ytter, & Fadiga, 2013; Green, Kochukhova, & Gredebäck, 2014), or explicitly: “fixating on the goal of an observed action before it is completed” (Elsner, Bakker, Rohlffing, & Gredebäck, 2014), “predict the outcome of the action before it is completed” (Keitel, Prinz, & Daum, 2014), “ability to predict ongoing events (e.g., looking at the final state of an event before accomplishment)” (Daum, Attig, Prinz, & Gredebäck, 2012). However, Flanagan and Johansson clearly were describing anticipation of the upcoming action in a sequence itself, not just its completion, and highlighted gaze coupling in space and time between actors and observers rather than merely proactive gaze arrival.

This conflation between modes of predictive gaze can lead to further confusion. For example, the authors of a recent paper took the “logical” next step and operationalized predictive gaze as gaze shifts from the hand to the target before arrival of the hand (Donaldson, Gurvich, Fielding, & Enticott, 2015). They then had to grapple with what to make of the original report on predictive gaze (Flanagan & Johansson, 2003) where “gaze was rarely directed to the hand in either performance or observation of hand actions”. Noting that “when viewing videos of simple grasping actions repeatedly, ‘predictive’ gazes might soon manifest as fixations directly to the target” (i.e. anticipatory predictive gaze), they then resolve the conflation by redefining the very notion of prediction itself “participants may have ‘known’ the outcome without the need for prediction” [my italics]. I would instead suggest distinguishing between separate modes of prediction: “participants may have ‘known’ the outcome [in advance] without the need for [online] prediction” (i.e. extrapolation).

Again, this conflation of “predictive gaze” with extrapolation needs to be understood against the background of a shift in the field away from investigating anticipatory predictive gaze as in the original
block-stacking design to investigating predictive gaze based on extrapolation on the basis of currently visible cues⁵.

5.2.3 Factors shaping modes of predictive gaze in action observation: concrete implications for studies

Next I will briefly revisit the organizational framework of factors which determine modes of predictive gaze to lay the groundwork for a systematic review of studies of gaze behavior in action observation (compare introduction 1.4.2).

5.2.3.1 When? Predictability of action onset, start targets, and inter-trial intervals

How predictable is the onset of a particular action to an observer? The timing of actions which are part of activity sequences may become transparent to an observer familiar with the activity, or when the action sequence is rhythmical and thus easy to read. The onset of the initial action is likely still unpredictable therefore introducing a start target before the target of interest into a design creates a predictable temporal context for subsequent actions.

In study designs based on repeated presentations of a single action, trial start cues and the temporal structure between trials will determine predictability of action onset. A variable delay at trial start before action onset will discourage anticipatory gaze shifts, as will extended inter-trial and trial start phase intervals. Presenting a fixation cross as a trial start cue controls gaze position but may lead to an artificial dissociation between overt and covert attention.

5.2.3.2 Where: target configuration and perspective

⁵ This is made explicit by Marsh et al. (2014): “Two action goals were used within this stimulus set so that the participant had to attend to features of the action in order to make a predictive saccade to the correct action goal. If only one goal was used, predictive saccades may occur without any attention to the actions.”
In the presence of a single uniquely defined target location observers will be more likely to generate either anticipatory predictive gaze shifts or, when the timing of action onset is not predictable, predictive gaze shifts based on extrapolation after a fairly short delay in response to action onset.

The presence of multiple potential targets will likely facilitate reactively initiated gaze shifts based on extrapolation of the action trajectory, unless the target configuration features a strategically located target which may draw gaze shifts based on a guessing strategy (even for reactively initiated gaze shifts, see chapter 4). Whether a given target region provides such a vantage point will depend on the observer's perspective on the scene.

5.2.3.3 How: relative movement duration

Action duration is a key determinant of how difficult it is for the observer to jump ahead to the target location in time for monitoring arrival and contact. When actions are performed in a short duration, i.e. for fast movements or movements to near targets, earlier gaze shifts are necessary to arrive in time with completion. Conversely, when observing actions which are executed slowly or to distant targets, the observer has much more time to monitor the unfolding of the action to extract information about the direction of movement. Thus, movement duration (distance & velocity profile) effectively sets the “difficulty level” of the task. Presenting very slow paced actions will likely facilitate tracking.

5.2.3.4 Why: action goals and task engagement between real actions and video recordings

We have proposed that effectively the same features which draw gaze for actors also are intrinsically interesting to observers. These include contact events, such as grasping or placing of an object and potential contact events like obstacles collisions which need to be avoided (Johansson, Westling, Bäckström, & Flanagan, 2001).

From our experience, real actions generally draw the gaze of observers much more robustly than video recordings, especially under free viewing conditions (in the absence of a defined task). Thus a wider
range of variability both within and between participants can be expected when video recordings are shown, including multiple modes of predictive gaze shifts as well as reactive gaze shifts.

5.2.3.5 Prior knowledge and skills on the part of the observer

Observers who are familiar with an activity may be able to “read” action targets and the unfolding of action sequences much better than naïve observers. This may translate into earlier and more accurate predictive gaze shifts both for anticipation and extrapolation. However, predictive gaze shifts in themselves are not a direct indicator of observer prior knowledge and skills – all the other factors listed here need to be taken into consideration as well. While there are a number of suggestive results demonstrating an important role of knowledge and skills on the part of observers in action observation, the nature of that knowledge and skills, in particular the notion that an action needs to be part of their “repertoire” for observers to engage in predictive gaze, still needs to be clarified, see section 5.4..

5.2.4 A systematic framework for comparing results across studies

Some of these factors are recognized in the literature, typically in terms of recourse to one of the studies on contextual modulation of gaze arrival times discussed below (section 5.2.6). However, recognition is uneven and the lack of an integrated view becomes evident when comparing results across different studies. For example, Ambrosini et al. (2011) compare mean gaze arrival times directly, noting that: “The magnitude of the gaze proactivity in our study is comparable with that found by Flanagan and Johansson (2003)” and then go on to discuss this difference in terms of the types of actions and motor cues involved. However, the authors overlook the actor's movement duration: In the Flanagan & Johansson (2003) study observers anticipated the onset of a fast paced action (duration of less than 0.5 s), whereas in their study observers typically reacted with a delay to a much slower paced action (duration of more than 1 s). Thus while gaze arrival relative to action completion happened to be similar between these two studies, this similarity is meaningless for most purposes other than for evaluating the visual consequences for the observer. Similarly, Rosander and von Hofsten compare “magnitude of gaze
advantage” - i.e. mean relative arrival time of hand and gaze – again between studies of highly different movement durations (Rosander & von Hofsten 2011). To explain these differences, they then compare angular distances between targets from the observer's perspective – which was especially pronounced in their study. However, the relative duration of an observer's gaze shift (typically 50-100 ms) is only a minor factor compared to the relative duration of the actor's movement (here on the order of 500-1000 ms), again pointing towards the need for a systematic perspective.

In a recent paper Gredebäck & Daum (2015) set out to construct a “model” of gaze behavior, discussing various challenges involved (identifying targets, disengaging from the hand, and so on) in terms of a “time line” of “component processes”. However, when the authors begin to put numbers to these processes, they characterize the time needed for adults to disengage from the hand as 38 ms again based on the original 2003 study on block stacking which featured anticipatory predictive gaze shifts. In contrast, infants' disengagement is characterized as 300 ms based on Rosander & von Hofsten's (2011) study which featured gaze shifts by extrapolation. The authors then conclude that “babies needed only 300 ms of movement information to disengage and accurately predict goals”. However, this conclusion is misleading on a number of levels. Infants needed on average 300 ms to respond to action onset whereas the single action goal was predictable all along, with particular infants being much faster (see discussion in section 5.4.3). A more meaningful comparison would be to contrast infant gaze onset times in Rosander & von Hofsten (2011) to those of the adult group in the same design, whose gaze latencies were highly similar to that of the infants. While there is a need for integrative frameworks in the literature, quantifying the duration necessary for gaze disengagement and target prediction of generic infants in generic situations may be itself questionable, given the number of operative contextual factors.

5.2.5 Constructing a hybrid methodology: Infants predict other people's action goals (Falck-Ytter et al. 2006)
Falck-Ytter and colleagues adapted the design of the original study (Flanagan & Johansson, 2003) and presented the first demonstration of predictive gaze in action observation in infants (Falck-Ytter, Gredeback, & von Hofsten, 2006). The authors added a number of changes to the experimental design which were picked up by psychologists studying infants as well as adults.

First, Falck-Ytter et al. switched to using videos rather than relying on a live demonstrator, which already brings their design much closer to established methodologies in developmental psychology (see Gredebaek, Johnson, & von Hofsten, 2009). The authors used three objects to be transported to a target location one by one akin to the original study on block stacking. However, Falck-Ytter et al. had the actor in the video clips place the objects into a bucket rather than on the table in full view of the observer, presumably so as to get rid of the object after transport so it cannot draw gaze again later on. A “happy face” was added to the bucket which moved in conjunction with a sound when one of the objects entered into it, presumably since actual placing events were now occluded and thus artificial events needed to be constructed to robustly draw the observer's gaze.

Second, the authors relied on areas of interest around the target location (bucket) and object start locations and used gaze relative to object arrival at the target area as their key measure. Their main results as reported in their paper are reproduced in figure 5.1a. In summary, adults and 12-month-old infants showed proactive gaze arrival (on average around 350ms in adults and 200ms for 12-month-olds), whereas 6-month-old infants did not (gaze arrival -200ms on average, i.e. after the hand).

Figure 5.1: a Mean relative gaze latencies at arrival for the adults and infants aged 12 and 6 months in the experimental condition (HA: human action) and control conditions (SP: self-propelled motion, MM: mechanical motion). b More detailed data from their supplementary table 2 showing relative gaze arrival times as a function of action number in the sequence. Note the shorter gaze latencies for the first movement for both adults and infants. Also note the diminishing number of valid gaze trials (N) presumably reflecting diminishing engagement over the course of watching the action sequence on video in both adults and infants. From Falck-Ytter et al. (2006).
While the original report had interpreted actor-observer gaze coupling in terms of the direct-matching hypothesis and noted parallels to the mirror neuron system, Falck-Ytter and colleagues went a step further and stipulated a “direct link between MNS [mirror neuron system] activity and proactive goal-directed eye movements during action observation”. The authors then cite the study by Flanagan and Johansson (2003) in support of the statement that “[i]n adults, the MNS is only activated when someone is seeing an agent perform actions, not when objects move alone” (even though the authors in this study . Accordingly, the occurrence of proactive gaze arrival in infants was taken as a test for the “MNS [mirror neuron system] hypothesis” whereas an alternative hypothesis – “teleological stance theory” – was operationalized as predicting no difference between gaze arrival between the human action condition and their two control conditions (self-propelled motion and mechanical motion).

The averaging of gaze arrival times over participants and conditions was therefore made necessary to conduct hypothesis testing using direct comparisons between group and condition means. However, while the authors state that “[p]reliminary analyses of the data distributions confirmed normality and homogeneity of variance”, this is hard to reconcile with their more detailed report of mean gaze arrival by action number in their supplemental materials (figure 5.1b). To actually interpret these gaze latencies – unfortunately again only reported in the aggregate for each number in the sequence – in terms of modes of predictive gaze, it is necessary to reconstruct gaze latencies at onset. As the movement durations for moving the 3 objects are given as “1.04, 1.47, and 1.07 s” (Falck-Ytter et al., 2006, supplementary material 4), it becomes clear that we are dealing with a bimodal distribution of gaze latencies at onset, with gaze initiation relative to the action onset being slower for the first action in the sequence compared to the second and third, in line with the findings of Flanagan & Johansson's (2003) block-stacking study (see section 5.2.3.1).

The 250ms gaze latency at arrival for adults for the first action suggests mostly reactively initiated gaze shifts. Gaze arrives on average around 150ms earlier for the subsequent actions in the sequence which could be due to an increased number of gaze shifts driven by anticipation, gaze shifts initiated earlier but still based on extrapolation, fewer pure tracking trials, or a combination of all of these factors.
For infants (at 12m), gaze to the first action within each of the video presentations arrived on average in time with the hand, suggesting a prevalence of tracking. Gaze arrival to the second and third action in the sequence occurs earlier by an impressive 300 ms on average, which is faster than what the adults showed for the first action (again, on average). This is likely based mostly on extrapolation, perhaps including tracking followed by extrapolation, though again it is hard to be sure from the data reported. Averaging the infant data clearly misrepresents actual arrival times, as figure 5.1b makes clear that the mean value of 200ms is outside of the standard error reported for each of the action numbers.

The authors are clearly aware of the latter as they mention the increase in gaze latencies between first to the second action in passing and confirm its statistical significance. The differences between gaze behavior reported (1) between adults and 12m-old infants compared to 6m-old infants, and (2) between the experimental condition and the control conditions are sufficiently robust that averaging their data – while setting a problematic precedent – does not invalidate the authors’ conclusions.

I have recounted the study by Falck-Ytter et al (2006) in some detail here as it has set the trend for many if not most subsequent studies on predictive gaze in action observation, in particular, but not limited to, infant studies. As shown below, many of the subsequent studies share the basic design based on playback of video clips, adding special effects to these videos, and the emphasis on arrival times and analysis on the aggregate rather than the single trial level (in line with using slow eye trackers, i.e. with sampling rates of 50 or 60Hz). This presents a case of “domestication” of the methodology, as data visualization and analysis techniques drawing from the motor control tradition are adapted to a style more familiar to (developmental) cognitive psychologists.

5.2.6 Studies showing an effect of contextual factors on gaze arrival times

I will here briefly discuss 4 studies which have been set up to demonstrate the effects of contextual factors in terms of the framework (see sections 1.4.2 and 5.2.3) and largely follow the methodology of Falck-Ytter et al. (2006) critically discussed above. When modulatory influences on observer gaze latencies in general are reviewed, these studies are typically cited in support of quite generic statements
such as: “Other factors that contribute to an infant’s ability to make predictive eye movements include prior visual experience with the observed events [Henrichs et al. 2014], saliency of the goal [Henrichs et al. 2012], and the effect of a distal goal [Gredebäck et al. 2009]” (Gredebäck & Daum, 2015). Here I will characterize these studies in a little more detail to critically discuss their scope and limitations (see table 5.2). For comparison, the original study (Flanagan & Johansson 2003) and its follow-up studies are characterized in table 5.1.

<table>
<thead>
<tr>
<th>Study</th>
<th>Question</th>
<th>Activity</th>
<th>Modes</th>
<th>When</th>
<th>How</th>
<th>Scene Config</th>
<th>Obs. Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flanagan &amp; Johansson '03</td>
<td>Actor &lt;-&gt; observer gaze</td>
<td>Live block-stacking</td>
<td>Anticipation (1st act: extrap.)</td>
<td>Rhythmic movement</td>
<td>Fast: 1 act/s dur: ~.5 s</td>
<td>familiar sequence</td>
<td>Intrinsic (real objects)</td>
</tr>
<tr>
<td>Rotman et al. (2006)</td>
<td>Predictability</td>
<td>Block lifting live demo</td>
<td>Anticipation Extrapolation</td>
<td>Start Block</td>
<td>Fast: reach dur: ~.5 s</td>
<td>2 targets: un/predictable</td>
<td>Intrinsic (real objects)</td>
</tr>
<tr>
<td>Reichelt et al. (chapter 4)</td>
<td>Predictability Configuration</td>
<td>Virtual game playback</td>
<td>Anticipation Extrapolation</td>
<td>Start Object</td>
<td>Fast: cursor move: &lt;.5 s</td>
<td>2 targets variable pos</td>
<td>Intrinsic (virtual)</td>
</tr>
</tbody>
</table>

Table 5.1: Overview of follow up studies to Flanagan & Johansson (2003). See discussion in introduction section 1.4.3. Notable features are highlighted in bold. [config: configuration; extrap: extrapolation; dur: duration]
<table>
<thead>
<tr>
<th>Study</th>
<th>Question</th>
<th>Activity</th>
<th>Modes</th>
<th>When</th>
<th>How</th>
<th>Scene</th>
<th>Obs. Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Falck-Ytter et al. 2006</td>
<td>Infant predictive gaze</td>
<td>Video object movement</td>
<td>Anticipation?</td>
<td>3 transport actions</td>
<td>Slow: move dur. ~1 s</td>
<td>Single Target (bucket)</td>
<td>Intrinsic (video)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Extrapolation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Eshuis et al. 2009      | End Effect Salience                     | Video object movement     | Extrapolation?       | 3 transport or jumps | Duration unreported | Single Target (bucket) | 1) Effects  
2) Silent |
|                         |                                         |                           | Tracking?            |               |                 |                     |                  |
| Gredebäck et al. 2009   | “Action Type”                           | Video of act sequence     | (Extrapolation)      | Action sequence | Slow: move dur. ~1.3 s | 5 objects right 5 locations left | Intrinsic (video) |
|                         | “Goal Type”                             |                           | Tracking/React       |               |                 |                     |                  |
| Henrichs et al. 2012    | Goal Salience                           | Video of object reach     | Extrapolation        | 800ms start delay | Slow: reach dur. >1 s | 1) 1 object  
2) 4-in-1 ♦ | intrinsic (video) |
|                         |                                         |                           | Tracking/React       |               |                 |                     |                  |
| Henrichs et al. 2014    | Goal Certainty                          | Video of object reach     | Extrapolation        | Start action with delay | Slow: reach dur. >1.5 s | 1) 3 obj same  
2) 3 obj var. | intrinsic (video) |
|                         |                                         |                           | Tracking/React       |               |                 |                     |                  |

Table 5.2: Overview of studies reporting effects of contextual factors: salience of end effects (added special effect sounds for contact events), goal salience (heap of 4 objects), goal certainty (same object vs variable object choice), and “action type” (reach/grasp and transport vs closed fist) and “goal type” (dropping in bucket vs. putting on table). Condition delimiters (1,2) and other notable features are highlighted in bold. [dur: duration; react: reactive gaze; var: variable; ♦: diamond shaped configuration]

The studies listed in table 5.2 are clearly modelled after Falck-Ytter et al. (2006), rather than Flanagan et al. (2003) as shown in the use of video stimuli, focus on extrapolation, and presentation of movements of longer durations. All these studies focus on infants except for the first one.

The first study “Predictive eye movements are driven by goals, not by the mirror neuron system” (Eshuis, Coventry, & Vulchanova, 2009) mounts an attack on Falck-Ytter et al. (2006) aimed at several of the weak points of their design. The authors replicated their study and created a novel control condition (“Tiddlywinks”) showing a finger flicking small plastic frogs into a bucket. Also they did note that the “end effects” (the special sound and motion effects that were added to an object entering the bucket, see
above) were conspicuously absent from one control condition (the mechanical motion condition) and ran all their conditions twice, with and without end effects. The authors largely replicate the findings of Falck-Ytter et al. (2006) in that observers show robust proactive gaze arrival only in the human action condition with end effects (around 200 ms which is a much smaller effect for reasons not explained) and in none of the control conditions. However, gaze arrived on average only in time with the hand in the absence of end effects. The authors exclusively focus on this result in their interpretation (and conveniently do not contrast the human action condition in the presence of end effects with the control conditions plus end effects as that would strengthen Falck-Ytter et al. 2006), and claim that their result invalidates the mirror neuron system hypothesis for predictive eye movements. To be fair, the practice of directly linking predictive gaze behavior to either the “mirror neuron system” or “teleological reasoning” as theories to be “rejected” or “confirmed” by a (single) study had been started by Falck-Ytter et al. (2006). This appeal to parsimony is clearly misplaced as these two theories are not incompatible and the neurophysiologists of the Parma group themselves do not consider motor resonance to be the only route to action understanding.

The effect of artificial enhancements (special end effects) as documented for this particular design raises questions about the use of (artificially enhanced) video stimuli generally. Beyond that, the generality of their interpretation is not warranted, as they did not actually address studies (such as Flanagan & Johansson 2003) that did not use videos let alone special effects.

Gredebäck and colleagues set out to study predictive gaze in more detail in their 2009 study on “action type and goal type modulate goal-directed gaze shifts in 14-month-old infants” (Gredebäck, Stasiewicz, Falck-Ytter, Rosander, & von Hofsten, 2009). The authors presented video clips of a sequence of 5

---

6 For example: “Actions that are not part of the motor repertoire of the observer and that therefore cannot be reproduced appear to be recognized in nonmotor terms. They are most likely understood based on visual description of the observed events and inferences of their consequences and/or goals” (Buccino et al., 2004).
movements. In the containment condition, the actor reaches for 5 objects on the right one at a time and deposits each in one of 5 small buckets on the left. In the displacement condition, the actor instead places them on the table. In a control condition, the actor only moved his fist along similar trajectories. The authors used a between subjects design with 2 groups of infants aged 10 months and 14 months. For 10 months-old infants, mean relative arrival times were consistent with reactive gaze behavior in all conditions (arriving ~200-300 ms after the hand). For 14 months-old infants, gaze arrived at the object in time with the hand in the displacement condition but only for reach actions and was reactive for object transport actions. In the containment condition, gaze arrival times were somewhat faster for both reach actions (proactive by 100 ms) and object transport actions (but still arriving on average after action completion). Mean relative gaze arrival followed the hand in the control condition.

These arrival times are somewhat delayed compared to the earlier study by Falck-Ytter et al. (2006), presumably reflecting the multiplicity of different movements and potential targets. To “account for the complexity of the design”, Gredebäck et al. decided to change their criterion of prediction by redefining proactive gaze arrival: “Latencies were considered predictive if average performance was faster than the adult reactive saccade latency of 200 ms”. They cite one of their own studies in support of this move (Gredebäck & von Hofsten, 2007), which, however, deals with reactions to unpredictable re-emergence events from occlusion rather than with extrapolating visible action trajectories.

While the authors interpret their result to show that “14-month-old infants incorporate information about both action type and goal type in their online assessment of others’ actions”, it needs to be kept in mind that the generic “action type” category actually refers to reaching and object transport actions compared to the non-functional action of moving a closed fist back and forth and the generic “goal type” refers to a slight difference in latencies for shifting gaze to landmarks (containers) as opposed to a an empty space on a table.

In the introduction, I have contrasted perspectives on gaze based on perceptual salience with those based on guiding of actions (section 1.2). These perspectives are combined in a study investigating how “[g]oal salience affects infants’ goal-directed gaze shifts” (Henrichs, Elsner, Elsner, & Gredebäck, 2012).
The authors presented video clips to infants (12m) showing an arm slowly reaching for either a single object, or to a configuration of four objects placed adjacent (forming a single large object, experiment 1) or close together in the shape of a diamond (experiment 2). The two experiments show essentially identical results: Mean gaze arrival for the single object condition was around 100 ms before the hand, presumably reflecting mainly tracking trials with some amount of extrapolation. Mean gaze arrival for the large/multiple object condition was significantly earlier with gaze arriving on average 400 ms before the hand. Even though they documented this robust modulatory effect of target number/size (or salience), the authors used a uncharacteristically large confidence interval of 95% around the mean so that proactive gaze arrival in the single object (low salience) condition was discounted as that interval (-49 to 192 ms) extended below 0 ms or concurrent arrival of gaze and hand. Thus gaze behavior in that condition did “not pass the threshold” and the quantitative effect becomes a qualitative one: “only infants in the high-salience condition were able to look at the goal ahead of time”. Note that one of the authors of this study who here employed a uncharacteristically strict criterion for predictive gaze arrival instead chose a much more forgiving definition to count gaze shifts arriving later than the hand as predictive, see Gredebäck et al. (2009) discussed above.

Whether or not their effect is qualitative or quantitative, the authors do report a robust influence of salience (in terms of size and number of objects). To interpret this finding, we need to look closer at the stimulus material used. The authors present a very slow moving hand at the top of the display, and have it rest motionless for almost one second before slowly moving on to the target (1160 ms for a distance spanning less than 10° of angular distance), thus inviting gaze tracking of the hand. The question now becomes at what point infants (12m) are willing to disengage from tracking of the hand and shift their gaze to the target (by extrapolation). It is this willingness to disengage that seems to be modulated by the size or number of objects at the target. It would be interesting to contrast this design with a Falck-Ytter et al. (2006) type situation e.g. with rhythmically repeated actions.

The final study discussed here shows a number of similarities to the one presented as chapter 4. The authors present 12-month old infants with videos of choice behavior, concluding that “goal certainty
modulates infants’ goal-directed gaze shifts” (Henrichs, Elsner, Elsner, Wilkinson, & Gredebäck, 2014). Infants were shown 12 presentations of a video clip showing a hand slowly reaching for one of 3 objects laid out along a line perpendicular to the reach. In the frequent condition, the arm always reached for the same object/location. In the non-frequent condition, target choice was random. The results are mainly reported in terms of mean gaze arrival, which led the hand in the frequent condition (around 200 ms for trial numbers 2-9), while gaze arrival tended to follow the hand in the non-frequent condition and in the first trial of the frequent condition (around -150 ms). Trials 10-12 are not discussed presumably due to increasing fussiness and loss of data.

Compared to chapter 4, the authors present only a cursory analysis of prediction and accuracy. Prediction rates – defined in terms of proactive arrival at any target before the arrival of the hand at the choice target – are reported as 0.62 for the frequent vs. 0.47 for the infrequent condition, a trend which did not reach significance. The accuracy rates (0.75 for the frequent, 0.47 for the infrequent condition) would warrant separate analysis of correct vs. incorrect guessing trials rather than pooling arrival times. Also the high .47 accuracy rate in the unpredictable infrequent condition is well above the expected rate of 0.33 for pure guessing and thus strongly indicates partially successful disambiguation of targets based on extrapolation of target trajectory. This points to a highly interesting achievement on the part of the infants which is not addressed in the article. Generally, it would be instructive to have more insight on gaze onset with regard to accuracy, target choice, as well as the consequences and relative arrival of corrective gaze shifts after incorrect guesses (compare figures 4.4, 4.6, and 4.7). The introduction of a start target (the hand in the video places a bowl on the table before reaching out) may have facilitated infants shifting their gaze away from the hand and in part may explain the infants’ quite impressive gaze behavior in this study, both in the frequent and in the infrequent condition. Frequency is used throughout the article by the authors and would appear to be a more conservative interpretation than goal certainty, as alternative interpretations could be raised in terms of goal familiarity or even re-inforcement learning.
5.2.7 Predictive gaze as reaction time: a problematic methodological analogy

In contrast to the studies discussed above - as well to most of the studies discussed below - the original report (Flanagan & Johansson, 2003) investigated anticipatory gaze shifts, and its follow-up studies (Flanagan, Rotman, Reichelt, & Johansson, 2013; Rotman et al., 2006, chapter 4, see table 5.1) mapped out the range of configurations and instructions which lead to anticipatory gaze shifts or facilitate prediction by extrapolation. This is a quite different approach than is taken by the majority of studies in the field at this point. In particular, gaze coupling between actor and observer – the original basis for the appeal to the direct matching hypothesis – is no longer present when observers instead reactively initiate gaze shifts. Instead, the alternative methodology is based on varying the properties of currently visible cues and measuring their effects in terms of gaze arrival times relative to action completion. These studies are set up to confirm that a given contextual effect has a significant influence in principle and are largely unconcerned about effect size or applicability to a larger range of situations and focus on mean gaze latencies as their basic unit of analysis. A sketch of what is captured – and what is lost – when observer gaze behavior is viewed from this perspective is shown in figure 5.2 by comparing gaze behavior to reaction times.
The main advantage of this methodology certainly is its remarkable simplicity – indeed, many papers include a “data reduction” heading in their methods section. As explained in the introduction (section 1.3.1), increased variability is to be expected when deviating from the prototypical design of the original block stacking study. Methods of dealing with this variability have been presented in chapter 4. To my

Figure 5.2 Analyzing and reporting predictive gaze in action observation in analogy to reaction time data. a Sketch of design, latency density plots, and mean latency bar graphs illustrating data analysis for 2 conditions/groups (red and blue traces) for a generic reaction time study. b Equivalent sketch for a study on observer gaze (black cursor) in action observation using regions of interest. Note the reliance on relative gaze to the arrival rather than the onset of the hand action. Were gaze data to be visualized in detail, it would readily become apparent that differences in mean gaze arrival times can reflect qualitatively different sources: fewer anticipatory gaze shifts, an overall delay in predictive gaze shifts based on extrapolation, more or prolonged episodes of tracking, or any combination thereof. Instead data are typically reported in terms of mean gaze latencies or as proportion of trials with proactive gaze arrival.

The main advantage of this methodology certainly is its remarkable simplicity – indeed, many papers include a “data reduction” heading in their methods section. As explained in the introduction (section 1.3.1), increased variability is to be expected when deviating from the prototypical design of the original block stacking study. Methods of dealing with this variability have been presented in chapter 4. To my
knowledge, the only other study which has even noted variability in gaze behavior in action observation explicitly is Gesierich et al. (2008). Given that almost all current studies use video stimuli (but see Rosander & von Hofsten, 2011) variability is likely a common challenge and the “black box” method outlined here at least offers a pragmatic solution.

Averaging out variability in the data may be premature, however, especially as all 3 modes of predictive gaze may lead to proactive gaze arrival. Taking the study reported in chapter 4 as an example, this would effectively mean relying almost entirely on the analysis shown in figure 4.5, and glossing over figures 4.4 and 4.7. This is highly unfortunate since many studies which follow this methodology report highly suggestive results (see examples above, table 5.2) which are hard to interpret further in the absence of more detailed analyses and data visualization.

A strong case can be made that gaze arrival at the target relative to the hand or cursor (“the standard measure used in the majority of eye-tracking studies with infants”, Henrichs et al., 2014) is the operative parameter to assess the consequence of gaze shifts in action observation functionally in terms of allowing monitoring of contact events with central vision (see chapter 4, figure 4.7). However, it seems inexplicable that this parameter is instead chosen as an indicator of prediction. This leaves open a number of ways to adjust the “difficulty setting” of the experimental task, in particular to fine-tune movement distance and speed, i.e. to shorten or prolong the duration of the demonstration in order to make it harder or easier for gaze to overtake the hand en route to the target. Clearly, gaze onset relative to action onset is a much cleaner measure of prediction, can actually be regarded as a reaction time, and has the added benefit of distinguishing between modes of prediction (anticipation and extrapolation).

Finally, the analogy to reaction time data would explain the treating of aggregate data by condition as the basic level of analysis prevalent in the literature which leads to essentially discarding trial by trial variations as noise. However, gaze initiation time has direct consequences on the subsequent visual input of the observer and thus needs to be interpreted in terms of observer choice on a per trial basis. A particular gaze shift can be said to arrive at a target before or after the hand, whereas gaze arrival
averaged over multiple trials and participants cannot meaningfully be said to be proactive or reactive when there is variability in the dataset.

The looking time paradigm (Aslin, 2007) provides another established blueprint for dealing with eye tracking data which could have further reinforced the preference for summary statistics. Conversely, the methodology of the original study and follow-up studies drew heavily from the motor control tradition, in particular analysis of eye-hand coordination, used to providing visualization of raw movement traces and analysis of trial by trial variation. Such tools may not be part and parcel of the standard approach of cognitive (developmental) psychologists and social neuroscientists interested in action observation.

5.3 The relationship between predictive gaze behavior in action observation and the mirror neuron system

5.3.1 Goal prediction, action understanding, and learning from observing

The mirror neuron system has been constantly appealed to ever since Flanagan and Johansson's original (2003) paper tied the close coupling between actor and observer gaze behavior in block stacking to the direct matching hypothesis (see Falck-Ytter et al. 2006, discussed in section 5.2.5 above). While such appeals may have increased the visibility of eye tracking studies, they may inadvertently have contributed to confusion about the predictive processes involved. In any case, designs based only on eye tracking cannot begin to clarify this relationship.

Goal prediction is at the heart of the account of the group around Giacomo Rizzolatti in Parma who settled on action understanding as the primary function of the mirror neuron system (Rizzolatti & Craighero, 2004; Rizzolatti & Sinigaglia, 2008; Rizzolatti & Fogassi, 2014). Their account is typically couched in terms of action goal prediction based on extrapolation of observed movement kinematics, including inferring target location, action type (i.e. precision vs. whole hand grip), and target object identity. However, this conceptual account remains somewhat vague and descriptive and certainly does
not begin to reflect the detailed findings about mirror neurons that this group has accumulated over two
decades of intensive study. At present there is no consensus about the actual nature and level of detail of
goal prediction contributed by the motor system in action observation nor about its role in action
understanding (see Gallese, Gernsbacher, Heyes, Hickok, & Iacoboni, 2011), nor is there a consensus
about how to characterize the mirror neuron circuit in computational terms (Fleischer, Caggiano, Thier, &

Alternative perspectives on mirror neuron functioning have been proposed to expand the emphasis on
“motor activity during observation of an ongoing movement”. Such a focus may “possibly [lead] to a
view of this mechanism as a passive, automatically triggered motor ‘echo’ used for action recognition”, as
opposed to a more active role for mirror neurons in “setting up an anticipatory model of another person’s
action” (Kilner, Vargas, Duval, Blakemore, & Sirigu, 2004). This has been argued forcefully in terms of
the predictive coding approach in computational neuroscience which regards generating future state
estimates based on integrating information from all sources as the primary function of brain mechanisms
in general. Applied to the mirror neuron system (Kilner, Friston, & Frith, 2007a, 2007b; Kilner & Frith,
2008; Kilner, 2011), this results in a bi-directional account of the system as both generating and receiving
state estimates (see Miall, 2003; Caligiore, Pezzulo, Miall, & Baldassarre, 2013, for a discussion of the
mirror neuron system in terms of forward and inverse models). Thus proponents of the predictive coding
perspective have been much more open about discussing the mirror neuron system in relation to
anticipating actions based on prior knowledge, which before had been raised mainly in connection with
“deflationary” accounts of the mirror system which relegate its role to spelling out the details of prior
beliefs leaving reasoning to be the driving force behind social cognition (Csibra, 2005, 2008).

As the predictive coding approach is primarily framed in terms of brain activity in the service of
generating estimates of future states – one could say all roads, while bidirectional, lead to prediction – its
proponents have been less forthcoming about the function(s) of sensory input (whether actual or
predicted) obtained through action observation. In contrast, Marc Jeannerod (1997, 2006) proposed
learning by observing as a more ambitious interpretation for mirror neuron function, which is generally in
line with our proposal that observers monitor actions and in particular contact events in order to keep track of and update estimates about the world in the service of guiding future actions. When motor learning is understood in terms of a comparison process between predicted and observed feedback (Wolpert & Doya, 2003), the two interpretations (action understanding and learning) begin to effectively align, different only in their emphasis on prediction per se as compared to prediction as a means to achieve effective monitoring (compare the discussion of Flanagan et al., 2013 section 1.4.3).

5.3.2 The mirror neuron system and predictive gaze based on extrapolation

While the predictive gaze in action observation and the mirror neuron firing characteristics are both object oriented and involve goal prediction, there is at least one clear difference: proactive shifting of gaze to a target is certainly not mandatory and intimately linked to attention, whereas mirror neuron firing has an automatic character and certainly does not require target fixation on the part of macaques who at times may not be very cooperative during the elaborate protocols of stimulus presentation.

To clarify this relationship, Monica Maranesi and colleagues recently conducted an experiment where eye movement data and F5 mirror neuron recordings were collected at the same time (Maranesi et al., 2013). Monkeys performed and watched a live experimenter perform self-initiated reaching, grasping, and lifting movements towards a small cube of food. In action observation, the firing rate of about half of F5 mirror neurons was consistently elevated in trials where monkeys shifted their gaze to the object (gaze-dependent) whereas the firing rates of the remaining half of recorded cells was unaffected (gaze-independent). Gaze-dependent mirror neurons also fired more strongly in proactive trials where the monkey's gaze arrived at the target before the experimenter's hand at the goal object compared to reactive trials when their gaze arrived later (there were very few anticipatory gaze shifts as action onset was initiated by the experimenter at variable intervals). Moreover, the activity peak of mirror neurons modulated by gaze was closely coupled to the timing of the eye movement relative to hand-target contact,

---

7 Intriguingly, training monkeys to fixate on the screen has proven instrumental in eliciting mirror neuron activation from video stimuli which for a long time had proven elusive (Caggiano et al., 2011).
suggesting an integrated eye and hand motor system predicting upcoming contact events which includes a portion of mirror neuron cell ensembles.

To date, this experiment constitutes the only direct demonstration of a relationship between the mirror neuron system and proactive gaze behavior. While accounting for the automaticity of mirror neurons on the one hand – which always responded to the experimenter's grasping action irrespective of attention – it is also suggestive of a substantial interconnectedness between motor systems for hand and gaze control in action observation as the spiking intensity and timing of about half of the mirror neurons recorded were modulated by gaze.

There have been a number of experiments aimed at linking proactive gaze to the mirror neuron system in indirect ways. I will here discuss their basic paradigm (Ambrosini et al. (2011) on which they based a series of studies (at least 7 to date). The authors designed their behavioral study to stay as close as possible to the concerns of the Parma account of the mirror neuron system for grasping actions, in particular the specificity of the mirror neuron circuit to whole-hand precision grip grasping (illustrated in figure 5.3).

![Figure 5.3 Video still illustrating the spatial configuration of the stimulus set used by Ambrosini et al. (2011) as well as in its numerous follow-up studies. Note the fixation cross on the hand at trial start which effectively rules out anticipatory predictive gaze shifts.](image)

Observers were presented with videos of a hand grasping either the large (whole hand grip) or the small object (precision grip). As a control condition, the authors interspersed trials where the actor is
reaching towards one of the objects with a closed fist so no grip aperture information is present ('no-shape” condition).

The authors report higher accuracy – first gaze shift to the object subsequently grasped – in the preshape condition (82%) than in the no-shape condition (65%), earlier mean gaze shift onsets in correct trials (193 ms compared to 221 ms), fewer saccades (1.62 compared to 2.06), and earlier mean arrival times in correct trials (-176 ms compared to -70 ms for large targets, -109 ms compared to +90 ms for small targets between preshape and control, respectively). Thus the difference in mean latency between the preshape and the control condition is much less pronounced at onset than at arrival. Since this is the case for correct trials, this does not reflect corrective saccades, but instead very likely reflects more frequent and/or prolonged tracking of the hand in the control condition, which triggers gaze onset immediately but leads to a delay in arrival.

It is very difficult to reconstruct how actual gaze behavior looks like in this study since the data are reported as averages over many trials and participants. Data analysis is performed in terms of ANOVAs so that main effects reflect a lumping together of different gaze strategies (guessing, tracking, reactive), with their respective onset times and consequences of those strategies. There was significant amount of guessing even in the preshape condition as we can expect 18% of incorrect guesses to be accompanied by at least an equal number of lucky hits, making it difficult to appreciate in detail how the preshape condition leads to somewhat more accurate and slightly faster arrival times overall. In addition, the control (no-shape) condition is itself problematic, as it is not matched in terms of prior familiarity and the closed fist may not indicate goal-directedness to the same extent as a grasping hand and facilitate tracking (Gredebäck, Stasiewicz, et al., 2009). Showing normal grasping directed at one of 2 objects of the same size would have been a more informative control condition.

The most innovative and at the same time most limiting aspect of this setup is the narrowness of its spatial configuration deliberately designed to meet several necessary conditions to study goal prediction based on grip aperture. First, there are two targets of quite different sizes. Second, the two targets are placed in close proximity, making it hard to disambiguate target choice based on movement trajectory
which observers are otherwise able to exploit quite early on in the movement, that is, well before preshape information becomes available (Rotman et al., 2006, also see chapter 4). Third, those targets are not actually adjacent, and the small gap between them is accentuated by placing the camera/observer very close to the action. Thus, gaze analysis based on areas of interest around the objects becomes possible and observers are discouraged from simply shifting their gaze towards their “center of gravity”. In fact, observers clearly fixated the large object much more readily which the authors interpret in terms of salience but which could equally well represent gaze shifts directed at both objects at once. In functional terms, observers looking at either object are already able to monitor deceleration and contact with high acuity irrespective of which object ends up as the action target so it is not clear what the benefits of the apparent capacity to use grip aperture for target prediction are for observers.

When interpreting studies based on this design, we need to keep in mind the sheer length to which Costantini and colleagues needed to go to make goal prediction based on pre-shaping tractable: creating a scene where two objects of different size are located close enough to each other to so as to prevent disambiguation by trajectory but just far apart from another to (at least eventually) merit individual fixation on the part of an observer close to the action. Accordingly, goal prediction based on pre-shaping has never been demonstrated to play any role whatsoever in naturalistic settings, rather, these studies were meant to provide indirect evidence for mirror neuron system involvement in predictive gaze. Unfortunately, these studies (there are at least 6 more by the Costantini group not discussed here) are all too often reviewed in generic terms which are actually misleading while technically correct: “observers take advantage of […] specific motor cues like hand pre-shaping to predict other people’s actions” (Elsner et al., 2013), in “complex scenarios, where more objects of varying shapes and sizes are present” (Causer, McCormick, & Holmes, 2013).

5.3.3 The mirror neuron system and anticipatory predictive gaze

While we are unfortunately still lacking a study directly exploring the interplay between anticipatory predictive gaze and the mirror neuron system, there have been recent developments in the study of mirror
neurons which are suggestive of such a link. In one of the landmark studies of the Parma group (Fogassi et al., 2005) monkeys watched an experimenter either grasp and eat a piece of food, or – marked by the presence of a container in the workspace – grasp and place the piece of food into that container. The firing rates of parietal mirror neurons for grasping were modulated by the overall intention behind this 2-step action, with most mirror neurons showing firing increases for grasping-to(-subsequently)-eat and some showing increases for grasping-to-(subsequently)-place. This study is crucial since it expanded the horizon of mirror neuron research, now also looking at how contextual knowledge informs the mirror neuron system rather than only focusing on how the mirror neuron system contributes to action understanding.

A recent fMRI study has applied this “bi-directional” perspective by using dynamic causal modeling (Gardner, Goulden, & Cross, 2015). Participants were shown sequences of initially novel dance movements and became familiar with them over the course of the experiment. BOLD network activity between the the action observation network and perceptual nodes developed in a way compatible with the notion that mirror neuron system activity is predominantly shaped by receiving perceptual input when actions are unfamiliar but action predictions contributed by the cortical motor system in turn start to drive brain activity in perceptual systems once actions are recognized as familiar.

The Bonini group has also begun to explicitly discuss the concerns of predictive coding in their recent experiments (Bonini, Maranesi, Livi, Fogassi, & Rizzolatti, 2014; Maranesi, Livi, Fogassi, Rizzolatti, & Bonini, 2014; see also the concise review in Urgen & Miller, 2015). In their two studies, monkeys either perform or observe an experimenter performing a go/no-go task. The authors report a sizable proportion of mirror neurons also firing in trials when the action needs to be inhibited. There are certainly multiple interpretations of the nature of the link between cue-association, mirror neuron function, and inhibition. For present purposes, the key result is the demonstration of “predictive” compared to “reactive” mirror neurons (Maranesi et al., 2014). Predictive mirror neurons began increasing their firing rate already in anticipation of movement onset (-340 ms) rather than in response to it as was their case in their prior study discussed above where the timing of movement onset was unpredictable (+60 ms, Maranesi et al.,
Predictive mirror neurons formed a minority (20%) of “action mirror neurons” and a majority of “inaction mirror neurons” (60%). Predictive mirror neurons were characterized in response to overlearned auditory association cues only. While we still can only speculate about their possible role during actual anticipatory gaze behavior for now, it appears likely that (subsets of) mirror neurons could indeed be active when watching activities which follow a predictable temporal structure such as block stacking, as suggested more than a decade ago by Flanagan and Johansson (2003).

5.4 Developing a capacity and inclination for predictive gaze behavior

5.4.1 Expertise and practice effects in action observation

Recent debates about the role of the motor (mirror) system in action understanding are often centered around the question of expertise – to what extent do people who are highly familiar with an activity have a “deeper” understanding of the action? In particular, can the role of the observer's own motor skills be disentangled from visual familiarity? There are a number of studies investigating the role of expertise and practice effects on action understanding, so I will only briefly mention a few particularly striking ones. In an elegant demonstration of the efficacy of motor training, Casile and Giese (2006) trained participants to perform an unusual walking coordination in the dark and demonstrated that recognition of point light displays of these movements was significantly improved after this non-visual training. Aglioti et al. (2008) showed enhanced ability to predict the outcome of free shots in basketball for experts compared to novices with combined motor and visual experts (athletes) outperforming mostly visual experts (coaches). The recent study by Gardner et al. (2015) discussed above merits mention for going beyond performance benefits of expertise and correlated increases in activation of parts of the observer's motor system to instead investigate changes in the interplay between motor and perceptual systems over the course of familiarization.
There appears to be a striking lack of studies investigating the effects of familiarity, let alone expertise, on gaze behavior in action observation. This is not surprising as observers in Flanagan and Johansson's (2003) block stacking study were essentially at ceiling performance throughout the experiment, and Falck-Ytter et al. (2006) study likewise did not show any learning effects for infants or adults. Investigating familiarity and practice effects would likely require novel designs.

5.4.2 Motor abilities as prerequisites for action observation: 3 caveats

In contrast to the absence of studies directly looking at practice or expertise differences in adults, expertise has arguably been the main focus of investigations into the development of predictive gaze in action observation. Beginning with the strong interpretation of the link between the mirror neuron system and predictive gaze in Falck-Ytter and colleagues (2006), a number of investigations have set out to show 1) a positive correlation between infants' relative gaze arrival times when watching a particular action and their own ability to engage in that action and 2) that infants begin to show proactive gaze arrival for a given action only just after they have started to engage in that action themselves. Before reviewing studies exploring the developmental trajectory of infant gaze behavior in action observation, three caveats are raised here.

First, the language of “actions” – really categories of actions – being “mastered” or “becoming part of the motor repertoire” quietly presupposes a systematic understanding of generalization of actions. To what extent can an observer who encounters tennis for the first time use their particular set of motor skills to make sense of a forehand stroke: To what extent can they bring experience in playing table-tennis to bear? Throwing objects? Hitting or “banging” behavior shown by infants? Conversely, to what extent does a child's experience of actually having held a racket and flailed at tennis balls contribute to her

8 A recent study has attempted to show practice effects for a virtual version of the block-stacking task (Möller, Zimmer, & Aschersleben, 2015). However, their study is confounded by the inclusion of “incorrect” actions – where the actor reaches for the wrong block – raising questions about predictability irrespective of practice.
making sense of an athlete’s performance of the “same” action – which may well turn out to be only superficially similar when analyzed in biomechanical and control theoretic terms? Generalization patterns will likely prove a lot more complicated, tenuous, and potentially counter-intuitive than conceptual similarities might suggest judging from investigations into the far better documented case of adult object manipulation: to what extent does a learned skill generalize to working in a different part of the workspace, with an outstretched arm, with a different grip on the object, with a different object, when hands are switches, etc. (Wolpert, Ghahramani, & Flanagan, 2001; Witney, 2004; Braun, Mehring, & Wolpert, 2010). Furthermore, when matching an action plan to an observed action we are necessarily dealing with a generalization process involving additional complexities, as proprioceptive input from the observed action is lacking, the action is seen from a different perspective than would be encountered in self-generated action, and movement styles and body proportions may differ substantially between actor and observer.

Second, as documented in detail in the previous sections, an adjustable “difficulty level” is set by the choice and properties of the action presented, as well as its surrounding context. A number of measures can be taken to facilitate early gaze arrival relative to completion of the action: increasing movement time, presenting a single prominent target object, making the movement appear more clearly goal-directed, repeating movements and making onset times more transparent, etc. Thus systematic differences between age groups may be partly due to these external factors rather than directly reflecting qualitative differences in infants' skill level.

Third, there is a close relationship and hence a principle confound between infants' capacities and inclination. This is particularly relevant as the striking motivation of infants to engage in novel actions at the edge of their abilities has even been proposed as a central motive in action development (von Hofsten, 2004, 2009) and the observer's motivation is an important factor in our proposed framework. Negative results are indeed likely to at least partially reflect less than stellar interest of infants for actions which they do not recognize as relevant to them particularly when shown as video recordings. We thus need to complement the investigation into actions which infants are just acquiring, and also look at actions of
caregivers which are of interest to infants in their current life situation irrespective of their own level of motor development. Indeed, very young infants often become intensely absorbed in watching their caregivers engage in quite complex everyday activities (Rossmanith, Costall, López, & Reddy, in preparation).

5.4.3 Developmental trajectories for engaging in predictive gaze in action observation

In this section I review selected studies which shed light on the development of predictive gaze in infancy, in particular its relationship to motor abilities. The studies are briefly summarized in table 5.3.

<table>
<thead>
<tr>
<th>Study</th>
<th>Question</th>
<th>Activity</th>
<th>Modes</th>
<th>When</th>
<th>How</th>
<th>Scene</th>
<th>Ages (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Falck-Ytter et al. 2006</td>
<td>Demonstration; correspondence</td>
<td>Reach-grasp transp vids</td>
<td>Anticipation? Extrapolation</td>
<td>3 acts</td>
<td>Slow: move dur. ~1 s</td>
<td>Single Target (bucket)</td>
<td>6,12, adults</td>
</tr>
<tr>
<td>Kanakogi et al. 2011</td>
<td>Developmental correspondence</td>
<td>Reach-grasp video Clips</td>
<td>Extrapolation Tracking</td>
<td>Single action</td>
<td>Very Slow: duration: 2 s</td>
<td>2 objects from above</td>
<td>4,6,8,10, adults</td>
</tr>
<tr>
<td>Rosander et al. 2011</td>
<td>Similarity btw. action/percept.</td>
<td>Live object placing</td>
<td>Extrapolation</td>
<td>Single action</td>
<td>Average: dur ~.8s</td>
<td>Ball in one hand tube in other</td>
<td>10-11 adults</td>
</tr>
<tr>
<td>Hunnius et al. 2010</td>
<td>Familiarity with actions</td>
<td>Naturalistic videos</td>
<td>Anticipation Extrapolation</td>
<td>Single action</td>
<td>Slow? Speed not reported</td>
<td>Frontal view of person</td>
<td>6,8,12,14, 16, adults</td>
</tr>
<tr>
<td>Green et al (submitted)</td>
<td>Cultural familiarity Chopsticks / Spoon vids</td>
<td>Extrapolation Tracking</td>
<td>Two step</td>
<td>Left→Right→Mouth; no info</td>
<td>Frontal view of person at table</td>
<td>8; sweden/china</td>
<td></td>
</tr>
<tr>
<td>Cannon et al. 2012</td>
<td>Familiar Object vs Location</td>
<td>Reach-grasp video clips</td>
<td>Extrapolation Tracking</td>
<td>Single action</td>
<td>Very Slow: dur 1.5s, +.5s in test</td>
<td>2 objects from above</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 5.3: Overview of selected studies on the development of predictive gaze in action observation. Key features are shown in bold, as are age groups (in months) who showed proactive gaze behavior. [dur: duration; percept: perception; transp: transport; vids: video clips]
Already the study by Falck-Ytter and colleagues (2006, see section 5.2.5), was conceptualized primarily in terms of a proposed developmental correspondence between abilities of action production and action observation. While their results – negative for 6 months old infants and positive for 12 months olds – are certainly generally in line with this notion as infant's “begin to master the action shown [placing objects in a bucket] at 7-9 months”, the large gap between the two age groups clearly does not warrant specific interpretations (but see Cannon, Woodward, Gredebäck, von Hofsten, & Turek, 2012).

More detail was provided by Kanakogi and Itakura (2011) who tested a wider range of infants (at 4, 6, 8, and 10 months of age), and also collected a simplified measure of infants' reaching ability in terms of the alignment between the two hands expressed as an angle in the range of 90° (both hands) to 180° (more “mature” one-handed reach). They reported a slight but significant correlation between this angle and mean relative gaze arrival for infants irrespective of age group. While their study is closely modeled after Falck-Ytter in framing, interpretation, and discussion of research question and results, the authors present reach and grasp actions as close-up clips of a hand and arm as seen from above reaching very slowly over a table towards one of two potential target objects. This design is derived almost directly from earlier studies using a habituation (looking-time) paradigm (Woodward, 1998), while the design of the Falck-Ytter study was adapted from Flanagan & Johansson (2003). The extended movement duration by itself may account for the positive result found in this study already for the group of 6 months old infants.

A very different approach was taken by Rosander and von Hofsten (Rosander & von Hofsten, 2011) who offer refreshingly detailed descriptions both of production and live observation of infant (10-11 months) and adult object transport actions. Participants either saw the experimenter move a small ball held in their left hand towards a transparent tube held in their right hand and insert it, or received the ball themselves at the same start position to place it into the tube held by the experimenter. Infants' gaze behavior was measured by means of EOG (electrooculography) which was integrated with 3D motion tracking of the head (which they could turn freely) and hand. All movements were performed in the horizontal direction as this technique does not allow accurate measuring of vertical gaze. While data from
individual participants is still averaged, the authors show example gaze and hand movement traces and provide a detailed treatment of gaze onset, arrival times, and number of saccades.

In action trials, hand and eye movements were initiated at around the same time on average for both adults and infants, the familiar pattern for eye-hand coordination. Gaze onset in action observation was significantly delayed relative to the experimenter's hand movement for both adults and infants (~200 ms for adults on average, ~300 ms for infants on average) but tended to arrive well before the hand at the target for both groups. Clearly, both infants and adults primarily engaged in predictive gaze based on extrapolation, very likely since the timing of the action onset was not transparent to the observers in the absence of a start target (a point whose significance was missed by the authors). This study demonstrates the feasibility of using live action demonstration in infant eye tracking studies and the quality of data collection and analysis matches well controlled studies of action observation in adults. We are still lacking data of comparable quality for anticipatory predictive gaze shifts as well as for gaze behavior of younger infants.

More insight into the development of predictive gaze by extrapolation is provided by a large-scale study of Hunnius and Bekkering (2010) who report gaze data from infants in age groups of 6, 8, 12, 14, and 16 months, as well as a separate adult control group. They showed three different video clips showing either a woman grasping a cup and bringing it to her mouth to drink, grasping a phone and bringing it to her ear, or grasping a brush and then bringing it to her hair (top of her head). “Nonfunctional” – and thus unfamiliar – uses of the same objects were shown in three additional control video clips, i.e. bringing the cup to the ear, bringing the phone to the top of the head, and bringing the brush towards the mouth. Unfortunately, no details are reported about the timing of action phases which is said to have been comparable between conditions. Data analysis is performed using regions of interest around goal locations (fairly small so as to be action-specific) and data are reported exclusively as proportion of proactive trials (i.e. earlier arriving earlier than the object at the action target). While the additional video clips showing actions to unfamiliar targets provide a useful baseline, their inclusion may have changed
the infants’ looking behavior. Goal locations are not well matched, i.e. the mouth is a more central and prominent target than the ear, which may account for some of the differences between conditions.

The results are striking (and likely unexpected judging from the choice of age groups): Already the youngest group tested (6 months) show positive results both in terms of the frequencies of trials and percentages of infants showing proactive gaze behavior – 5-10% for brush to hair, 10-15% for phone to ear, 25% for cup to mouth – with older infants showing only quite minor increases. Some of these scores are significantly higher than those for the matched control non-functional/unfamiliar videos. While interpretation is difficult in the absence of more detailed results, these data suggest that (some) infants do show proactive gaze behavior for familiar everyday actions which they themselves cannot yet engage in, or at least not in full-fledged form. The results for the unfamiliar actions suggest that infants in a few trials managed to predict the atypical action goal by extrapolation (although far less often than adults) and that they sometimes stick to the typical action goal (e.g. hair in the presence of a brush) even when this expectation is violated by the video clip (again although far less often than adults). The latter result may indicate predictive gaze by anticipation. The authors demonstrate that it is possible to get results on the basis of video clips of naturalistic actions without the need for special effects or very slow movement presentation, although the authors admittedly ran an impressive number of infants in their study (n = 283).

Additional evidence for a role of visual familiarity in enabling infants engaging in predictive gaze was provided by an elegant cross-cultural study by Green and colleagues (Green, Li, Lockman, & Gredebäck, submitted). The authors showed similar videos of a person bringing either a spoon first to a bowl of soup and then to the mouth or a pair of chopsticks to a bowl of noodles and then to the mouth. The authors ran two groups of 8 months old infants, one from Sweden and the other from China. While gaze arrival times at the bowl were usually reactive, arrival times at the mouth of the spoon were proactive on average for Swedish infants (~350 ms), but reactive when chopsticks were used. This pattern was reversed for Chinese infants who showed a clear advantage for chopsticks (~300 ms) but only a just positive mean result when watching video clips where the actor used a spoon. The latter result is a little unexpected – given that one would assume Chinese infants to also have come across spoons and spoon-feeding – but at
least indicates that the second group was not merely generally more adept at observing. This study provides some welcome nuance as the results are highly suggestive of a benefit of visual familiarity (growing up immersed in a culture where chopsticks are used regularly), while at the same time the authors also emphasize that this benefit was only present for eating actions which infants have substantial motor experience with – though certainly not via using chopsticks which is far beyond their fine motor abilities.

Recently, “predictive gaze” has begun to be used more widely in developmental psychology as a tool to address questions about infants' expectations more generally (Gredebäck, Johnson, et al., 2009). As part of this trend, the highly influential design of Woodward (1998) was adapted for use in an eye tracking study aimed at elucidating the cognitive processes that infants' goal prediction is based on (Cannon & Woodward, 2012). The authors presented carefully controlled videos showing either a hand or a claw very slowly reaching first along the midline and only later curving towards and grasping one of two objects (green plastic frog or red ball). In test trials, the two object were swapped and presentation stopped just before the movement trajectory revealed any information about the target location. The authors replicated their earlier finding showing that infants tend to expect the hand to continue to approach the same object as before whereas they expect a mechanical claw to move to the same location as before – but could now show that these expectations already manifested in predictive gaze shifts towards the same object and location, respectively.

The tradition represented by Cannon and Woodward (2012) brings an impressive depth of experience reflected in their level of experimental control of familiarity and habituation to objects and trajectories which go well beyond studies such as Falck-Ytter et al. (2006). Drawing on the field's detailed knowledge about to what extent visual stimuli appear animate and goal-directed to infants drawing from a range of careful studies primarily relying on the habituation-dishabituation (looking time) paradigm some researchers have called into question the “mirror neuron system theory”, instead proposing an alternative interpretation in terms of stimulus novelty and failure to disengage from mechanical claws, etc. (Southgate, 2013). The group around Amanda Woodward takes a much more conciliatory approach
cautioning against “using developmental arguments to minimize the potential significance of MNs [mirror neurons] for social cognition”. Instead, “the field should be pushing forward to understand the links between neural systems, social cognition, and motor skill” (Krogh-Jespersen, Filippi, & Woodward, 2014).

While a clearer exposition of what is needed would be hard to find, doing so will first and foremost require relating different research traditions with their different sets of concerns, concepts of prediction, and methodologies. This is quite evident from Cannon & Woodard's (2012) study itself: while their superior experimental control of stimulus novelty has already been noted, rather than submitting their data to automated visualization and analysis routines, the authors instead export videos of the scene with gaze position overlaid. These videos are then analyzed frame by frame by human coders in much the same way as they are in manual gaze coding in looking time studies. In the process the sampling rate is cut down from 50 Hz to 12 Hz. While adequate for their purposes, these relative standards showcase basic differences in orientation between social cognition researchers on the one hand, who view action observation primarily in terms of cognitive processes such as goal attribution and novelty detection conceptualized as taking place over time scales of (fraction of) seconds, and researchers in motor control on the other, who conceptualize observers as geared towards monitoring and updating state estimates about actors and objects by means of gaze shifts to be tracked on time scales of milli-seconds. Thus it is far from clear how to even begin to relate the notion of “abstract relational structure that organizes goal-directed actions” (Krogh-Jespersen et al., 2014) to e.g. Land's notion of a “knowledge base of the oculomotor system” (Land & Furneaux, 1997).

Apart from results based on eye tracking, there is an increasing number of studies based on a characteristic suppression in the alpha band (or more specifically the mu band) in the EEG likewise charting parallel courses of developmental trajectories for action and action observation (see e.g. Bakker, Daum, Handl, & Gredebäck, 2015; Cannon et al., 2015). Mu suppression is shown in action production and can be used as a marker for motor cortex involvement in action observation. One such study by Southgate and colleagues (2009), essentially an infant version of the design of Kilner et al. (2004),
suggests increased activity also in infants' cortical motor system prior to predictable action onset, thus at least partially addressing the gap left by the absence of developmental studies into anticipatory predictive gaze. However, the three caveats likewise apply to studies based on this methodology.

5.4.4 Conclusions

In general, the practice of presenting studies as testing the concept of developmental correspondence itself is beginning to lose its usefulness. Appealing to vague outside forces (mirror neuron system, general reasoning abilities) at this point rather serves to cut short actual discussion on interpretation of results and introduce confirmation bias as research focuses on linking gaze behavior in action observation of infants around the age that the infants begin to show the action studied rather than linking action and action observation abilities more generally. As discussed above, some recent studies have begun to move beyond this framing and the field is better served to follow their lead and focus on how particular findings may fit into a larger framework. Doing so will require elucidating how infants' skills for action and action observation – including but not limited to (predictive) gaze behavior – generalize to a range of (ideally somewhat naturalistic) contexts and to attempt to trace developmental trajectories of these skills (Karmiloff-Smith, 2010; Smith & Thelen, 2003).

5.5 Conclusion and Outlook

5.5.1 Gaze behavior as interaction

Gaze shifts have immediate visual consequences at onset, over their duration, and at arrival. At initiation they incur a cessation of monitoring the scene, in particular the current focal target. During execution they constitute a substantial transient loss of visual input. At arrival, they (may) unlock high

---

9 Presenting research results as talking points in a grand Manichean struggle about the ultimate nature of action observation between action and reason is of course still rather tempting, and enterprising researchers do not need to restrict themselves to any one side – compare two recent dissertations coming out of the same lab (Elsner, 2015; Green, 2014).
acuity information about the (estimated) next phase of the monitored action sequence. Therefore, seen as decision process, a gaze shift constitutes a complex gamble of costs and (estimated) benefits which need to be appreciated in visuo-motor terms. The ability to generate accurate predictions – or at least educated guesses – about the upcoming action target location based on generalizations of characteristics of objects, activities, and actors clearly is essential to effectively engage in action observation. However, knowledge acquired over the course of engaging in and observing activities, however such knowledge may be structured, will not necessarily be reflected in gaze behavior in a straightforward way (see chapter 4).

Predictive gaze in action observation thus constitutes a basic mode of environmental interaction, and does not merely reflect predictive processes but is instrumental in shaping visual exploration and thus social engagement itself. This fundamental point has been made a long time ago (Dewey, 1896; see Cisek, 1999; Cisek & Turgeon, 1999; and Cisek & Kalaska, 2010 for updated versions) and is now beginning to have more mainstream recognition (see e.g. Krauzlis, Bollimunta, Arcizet, & Wang, 2014).

Multiple research traditions can contribute to achieving a more integrated view on action observation from programming of gaze shifts to understanding socio-cultural activities. The present work aims to contribute to this project conceptually by proposing social (motor) learning as a key function of predictive gaze, by providing an organizational framework of distinct modes of factors which shape predictive gaze, as well as methodologically by showing how datasets which contain high variability can be analyzed with relatively simple tools to reveal causes and consequences of predictive gaze shifts.

5.5.2 Social Learning in Action between Motor Control and Social Neuroscience

A role for the motor system in action observation has first been suggested within neuroscience by the demonstrations of the visual “mirror” properties of neurons in F5 premotor cortex (reviewed in Rizzolatti & Craighero, 2004; Rizzolatti & Sinigaglia, 2008), and later also in parietal cortex (Fogassi et al., 2005) and primary motor cortex (Tkach, Reimer, & Hatsopoulos, 2007) – discussed here in relation to predictive gaze, see section 5.3. The close correspondence between observer and actor gaze in action observation first shown by Flanagan & Johansson (2003) further underlines that a motor control approach is necessary
to complement the visual perception perspective which has been typically used to study action perception before the discovery of mirror neurons (compare Giese & Poggio, 2003; with Casile & Giese, 2006; Blake & Shiffrar, 2007).

Here action observation seen as an active process of engagement with people and objects, further emphasizing the relevance of action planning and control already for observing. As such, the present work can be seen more generally as a small contribution towards a science of social interaction grounded in motor control. Such a field does not really exist yet, at least not in any organized fashion, and thus is still quite tractable and quickly reviewed: Psychologists interested in the interplay between action and perception have begun looking into joint action (Sebanz, Bekkering, & Knoblich, 2006), and social neuroscientists have launched the sub-field of interactive neuroscience (Hari & Kujala, 2009; Schilbach et al., 2013). Researchers in motor control have extended their concepts and paradigms to address action coordination (Noy, Dekel, & Alon, 2011; Dumas, Guzman, Tognoli, & Kelso, 2014), action understanding (Becchio, Manera, & Castiello, 2012), and social motor learning in reaching adaptation (Mattar & Gribble, 2005; Malfait et al., 2010). There are almost no studies on gaze behavior in social interactions (but see Noris, Nadel, Barker, Hadjikhani, & Billard, 2012; Yu & Smith, 2013).

At this stage these forays are quite disparate and are hard to relate as they reflect distinct concerns, concepts, and methodologies centered around different research traditions and experimental model systems. At the same time, we clearly have only begun to scratch the surface of the behavioral richness and flexibility of social interaction in cultural environments as revealed e.g. by researchers in situated action (Goodwin, 2013). More studies (and indeed research programs) are needed, as are organizing frameworks to systematically relate existing work.

5.6 References


