2017

Source memory for actions

Lange, Nicholas

http://hdl.handle.net/10026.1/8562

http://dx.doi.org/10.24382/515
University of Plymouth

Source Memory For Actions

PSYCHOLOGY
WITH
PLYMOUTH
UNIVERSITY

Nicholas Lange

School of Psychology

Plymouth University

Ph.D.

September 2016

Acknowledgements

Thank you to Timothy J. Hollins and Patric Bach for the supervision and continuous support throughout the last 3+ years. Thanks for the encouragement, and for the discussions about actions and memory frameworks and how they may combine.

Thank you to Laura, Dries and the rest of the PhDs who were around for most of this, for the community and the conversations in the pub. Thanks to Klaudia and Cee halfway around for the world for some perspective on life outside academia. Thank you to Charlotte for making this a great three years all around (and thanks to Sue and Tom for the lunches, dinners and the sailing).

Thank you to my parents for the support, the Sunday phone call check-ins and the visits. Thanks for being there whenever I come up with some new, surprising direction to take my life.

Abstract

Nicholas Lange

Source memory for actions

This thesis investigates source memory for performed and observed actions in recall and recognition tasks. The motor simulation account predicts that motor activation during action observation results in source misattributions of observed actions as self-performed. Alternatively, source judgements at test may be based on the evaluation of source features (source monitoring framework) or memory strength (relative strength account).

Experiments 1, 2 and 3 in Chapter 2 test if the motor simulation account explains false memories of self-performance after observation. Interfering with participants' ability to encode the motor trace during observation does not reduce participants' propensity to falsely recall observed actions as performed, but increases it.

Experiment 4 in Chapter 3 manipulates motor and visual interference at retrieval. Participants' false recognition of observed actions as performed and performed actions as observed is not significantly affected by motor or visual interference.

Experiments 5, 6 and 7 in Chapter 4 test if participants are better able to discriminate performed and observed actions if they generate the idea for the action they perform themselves. Participants' source discrimination in a recall task improves if they generate the ideas for self-performed actions (Experiment 5 and 6), only if they do not also generate ideas for actions they observe (Experiment 7).

Experiment 8 in Chapter 5 manipulates participants' visual perspective of actions they observe. There is no evidence for a significant effect of visual perspective during observation on subsequent false memories of self-performance in a recognition paradigm.

In my thesis I find no substantial support for a motor simulation account. While the results are broadly compatible with the source monitoring framework, model-based analyses show that participants' performance may be based on items' overall strength, in line with the relative strength account, rather than evaluation of source features.

Author's declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

Work submitted for this research degree at Plymouth University has not formed part of any other degree either at Plymouth University or at another establishment.

Relevant scientific seminars and conferences were regularly attended at which work was often presented, several papers were prepared for publication.

Word count for the main body of this thesis: **70700**

Signed: ________________________

Date: ________________________

Publications

Lange, N., Hollins, T.J, & Bach, B. (in prep) Disputes about everyday memories of actions.

Lange, N., Hollins, T.J, & Bach, B. (in prep) Generation effects in source memory for actions after enactment.

Lange, N., Hollins, T.J, & Bach, B. (submitted). Source errors in free recall of action memory. *under review in Frontiers*.

Posters and conference presentations

Lange, N., Hollins, T.J., & Bach, P. (2016, July). *Generation effects in source memory for actions after enactment.* Paper to be presented at ICOM 2016, Budapest, Hungary.

Lange, N., Hollins, T.J., & Bach, P. (2015, June). *No, you did that! Source errors in action memory.* Paper presented at SARMAC XI, Victoria, B.C., Canada.

Lange, N., Hollins, T.J.,& Bach, P. (2015, January). *(A source monitoring account of) Source errors in action memory.* Paper presented at the meeting of the Experimental Psychology Society, London, UK.

Lange, N., Hollins, T.J.,& Bach, P. (2014, April). *Memory for Actions: A two-way mirror?* Poster presented at the meeting of the Experimental Psychology Society, Kent, UK.

Contents

List of Figures

Chapter 1

Introduction

We generally have vivid memories of events we experienced. We remember how we broke our arm as we fell off the bike, or how we watched our sibling fall face-first into their birthday cake. Some people have experienced telling the story of such an event at a family gathering, only to have the sibling pipe up and say that it was them, not the speaker, who that event had happened to. Or, that they remembered that the speaker had fallen face-first into the birthday cake and not them. Sheen, Kemp, and Rubin (2001) reported that up to 70% of twin pairs reported at least one such disputed memory, while up to 20% of non-twin sibling pairs reported disputes about whether they or their sibling had experienced a particular event (Ikier, Tekcan, Gülgöz, & Küntay, 2003). One could argue that autobiographical childhood memories are prone to distortion due to the time that has passed since the event and the frequent retellings of an experience in that time. Yet confusions over whether someone performed and action or watched someone else perform an action can be observed in the laboratory as well, with up to 20% of observed actions falsely identified as self-performed (Hornstein & Mulligan, 2004). In this thesis, I will investigate how people confuse whether they performed or observed actions, if there are ways of reducing that confusion, and which theoretical perspectives can account for it.

In this introduction, I will first discuss people's ability to remember actions they performed in general by discussing the literature on the enactment effect. Following that, I will review the existing literature on source memory for actions and theoretical accounts that may explain the findings. Finally, I will outline the present work of this thesis.

1.1 The enactment effect

Independently of one another, R. L. Cohen (1981), Saltz and Donnenwerth-Nolan (1981) and Engelkamp and Krumnacker (1980) conducted experiments that came to the same conclusion: participants remember more items if they enacted the items at encoding.

Engelkamp and Krumnacker asked participants to encode simple items such as ‘Touch your right ear with your left hand’ or ‘Pick up the toy car’. All participants heard the action phrase being read to them. Some participants were instructed to merely listen to the verbal description, some to enact the verbal description or imagine enacting the verbal description. After a brief retention interval, participants were asked to verbally freely recall the items they had just encoded. Participants who had encoded items by enactment recalled more of them (.62) than participants who had only heard (.45) or imagined (.53) these actions.

R. L. Cohen (1981) asked participants to merely listen to verbal descriptions of simple actions such as ‘Wave’ and ‘Pick up the cup’ or encode them by enactment. After a brief retention interval, participants verbally freely recalled a higher number of items if they had enacted them (.55) than they did if they had only heard them (.44).

Saltz and Donnenwerth-Nolan (1981) showed participants action sentences such as ‘The bride stirred the coffee’. Participants in the enactment condition were asked to enact the action in the sentence (‘stir the coffee’), while participants in the visualization condition were asked to visualize the sentence. At retrieval participants were cued with the subject of the sentence (‘the bride’) and asked to recall the verb and/or object (‘stir the coffee’). Participants who had learned the sentences by enactment (.54) or by visualization (.67) recalled more verbs and/or objects than participants who encoded the sentences verbally (.36).

Participants’ superior recall was termed the enactment effect. Since then enactment, and the mechanisms underlying it, was explored using a number of experimental methods and action stimuli (for reviews see Engelkamp, 1998; Nilsson, 2000; Roediger III & Zaromb, 2010; Zimmer et al., 2001). Some of the types of stimuli that have been used are listed in Table 1.1. A bulk of the work on the enactment effect used a mixture of actions naturally performed without objects such as ‘Touch your left ear’ and performed with objects such as ‘Lift the toy car’, with real objects present (e.g. Bäckman, Nilsson, & Chalom, 1986; G. Cohen & Faulkner, 1989; R. L. Cohen, 1981, 1983; Kormi-Nouri, 2000) or imaginary (e.g. R. L. Cohen, 1981; Engelkamp & Krumnacker, 1980; Engelkamp, Zimmer, Mohr, & Sellen, 1994; Kormi-Nouri, 2000; Nyberg et al., 2001; Saltz & Donnenwerth-Nolan, 1981; Zimmer & Engelkamp, 1999). Aside from those actions, the enactment effect has also been investigated for dance movements (Foley, Bouffard, Raag, & DiSanto-Rose, 1991), sign language (von Essen & Nilsson, 2003; Zimmer & Engelkamp, 2003), sequences of body movements (e.g. Helstrup, 2001), sequences of action steps (e.g. Steffens, 2007) and applied real-life scenarios such as navigating while driving a car (Von Stülpnagel & Steffens, 2012).

Arguably, enactment and verbal learning differ in a number of ways, such as the richness of the experience, the motor activation and the presence or absence of objects. Object cues, for example, such as the body parts or objects used to perform an action boost memory after verbal encoding. When object cues are present and salient, the advantage of enacting items at encoded on memory retrieval can be severely limited or even eradicated (e.g. Engelkamp & Zimmer, 1997; Kormi-Nouri, 2000; Steffens, Buchner, & Wender, 2003; Steffens, Buchner, Wender, & Decker, 2007). While the enactment benefit occurs after motor activation at encoding, the motoric activation has to be uniquely coupled with a stimulus. Schaaf (1988) and Zimmer and Engelkamp (2003) asked participants to perform meaningless movements during encoding of verbal items. They did not observe enactment effects. Importantly, typically one action like ‘Lift the car’ is associated with one item. Here, the same meaningless motor actions were performed for all items. This did not allow for unique identification of an item by its motor trace.

Rather than just exploring the advantage of enactment relative to verbal encoding, much of the subsequent research has focused on comparing encoding by enactment to encoding by observation. Golly-Haring and Engelkamp (2003) (Experiment 1) asked participants to enact or observe simple action phrases like ‘Light a match’ using objects. In the test phase immediately following encoding, they observed higher corrected recognition (hit rates minus false alarm rates) for enacted (.95) compared to observed (.88) events (also see Engelkamp & Dehn, 2000). While Golly-Haring and Engelkamp tested recognition memory following recall, a similar memory advantage of enactment over observation was observed when recognition tests were the sole memory test. Mulligan and Hornstein (2003) asked participants to take turns enacting and observing actions. They reported correct recognition rates of .84 for enacted and .67 for observed actions after two days’ delay to test. Similarly, Hornstein and Mulligan (2004) reported corrected recognition rates of .74 and .63 for enacted and observed actions respectively after two days’ delay to test (also see Koriat, Ben-Zur, & Druch, 1991; Manzi & Nigro, 2008). This suggests a clear memorial advantage for actions after enactment relative to observation in a recognition task.

Table 1.1. Methods/Actions used to investigate the enactment effect

Description	Example	Example papers
Actions with one object	‘Lift the toy car’	e.g. Brandt, Bergström, Buda, Henson, and Simons (2013); R. L. Cohen (1981); Hornstein and Mulligan (2004); Lindner, Echterhoff, Davidson, and Brand (2010); McDaniel, Lyle, Butler, and Dornburg (2008)
Actions with objects in an array	‘Put the car next to the cup’	G. Cohen and Faulkner (1989)
Pantomimed action phrases	‘Lift the toy car’	e.g. Engelkamp et al. (1994); Leynes and Kakadia (2013)
Body movements on a grid		Helstrup (2005)
Body movements in space	Dance movements	Foley et al. (1991); Smyth, Pearson, and Pendleton (1988)
Collaborative creativity tasks	Making a collage	Foley, Passalacqua, and Ratner (1993); Hashtroudi, Johnson, and Chrosniak (1990); Rosa and Gutchess (2011)
Tracing/Drawing pictures		Helstrup (2005) ; Ratner, Foley, and Gimpert (2000)
Applied real-life scenarios	Navigating a car	Seamon, Philbin, and Harrison (2006); Von Stülpnagel and Steffens (2012)
Action sequences	Packing a backpack	e.g. Schult, von Stülpnagel, and Steffens (2014); Steffens et al. (2007); Steffens (2007); Steffens, von Stülpnagel, and Schult (2015)
Sign language		Zimmer and Engelkamp (2003); von Essen and Nilsson (2003)

In free recall tasks, the memorial advantage is dependent on the experimental design. Engelkamp and Dehn (2000) asked participants to encode actions by enactment and/or observation. They tested a pure enactment, a pure observation and a mixed enactment and observation group. The action phrases were actions like ‘Beat the egg’ that were pantomimed without objects. Participants’ performance at recall depended on both list length and between/within manipulation. The mixed group recalled a higher proportion of enacted actions than observed actions with short lists (.57 and .52 for enacted and observed actions respectively) and long lists (.41 and .35). For the between groups, this enactment effect was evident in long lists (.39 and .33 for enacted and observed actions respectively) but not for short lists (.53 and .60 for enacted and observed actions respectively) (also see Engelkamp & Zimmer, 1997; Golly-Haring & Engelkamp, 2003). Helstrup (2001) asked participants to either encode a spatial path on a grid by enactment or by observation. In an immediate recall test, participants were better able to recall the path after observation than after enactment (also see Helstrup, 2005). Similarly, Steffens (2007) asked participants to encode action sequences consisting of multiple steps (such as, individual steps to packing a backpack) by enactment or observation. She did not observe a significant enactment effect (for similar results see Schult et al., 2014; Steffens et al., 2015). This suggests that enactment in free recall typically leads to superior memory for actions only if actions were encoded in within-list or within-subject paradigms (for a discussion of possible reasons see Steffens et al., 2015).

1.1.1 The role of the motor trace

With some variation induced by experimental design discussed above, enacting an item at encoding leads to superior memory for this item at retrieval relative to verbal encoding or encoding by observation or imagination. Much of the research focused on exploring the crucial mechanism underlying this benefit in memory performance.

Engelkamp and Krumnacker (1980) proposed a fully multi-modal account (for a full description of their account see Engelkamp, 2001). Here the verbal information, the visual information and the motor information encoded during enactment are all encoded independently of one another (extending the dual-code theory (Paivio, 1971) by the motor code). They argue that the enactment advantage of enacted over observed actions, for example, provides evidences for the motor trace in particular (rather than the visual or verbal trace) improving the processing of the item information and thereby increasing memory performance for enacted actions. The predictions that follow from this account is that motor interference (but not other interference) should disrupt memory after enactment and that re-enacting actions at retrieval should increase the enactment effect.

The evidence in favour of motor activation specifically leading to the enactment effect are mixed. Motor interference immediately following encoding of target action sentences selectively reduces memory recall when sentences are enacted, but does not when they are verbally encoded or visualized (Saltz & Donnenwerth-Nolan, 1981; Zimmer & Engelkamp, 1985). In contrast, Helstrup (2001) did not observe a reduction of the enactment effect for action sequences after that type of retroactive interference. While he did observe reduction of the enactment effect under concurrent motor interference, the reduction was more severe under verbal interference. Similarly, both verbal interference and visual interference at encoding reduce the enactment effect for memory of action phrases (Bäckman, Nilsson, & Kormi-Nouri, 1993; Kormi-Nouri, Nyberg, & Nilsson, 1994), though note that neither study tested the effect of motor interference. In sum, while motor interference may impair the enactment effect, so do other kinds of interference.

Neuroimagining research suggests that motor information is reactivated during the verbal recognition of previously performed actions (Eschen et al., 2007; Masumoto et al., 2006; Nyberg et al., 2001), with some researchers reporting that the activation during the retrieval of imagined and observed actions is qualitatively similar but quantitatively lower than for performed actions (Senkfor, 2008; Senkfor, Petten, & Kutas, 2002; Wutte, Glasauer, Jahn, & Flanagin, 2012). This does suggest that motor information is retrieved when giving recognition judgements for performed actions. Behaviourally, Engelkamp et al. (1994) reported an increased memory benefit in recognition of previously performed actions when actions are re-enacted at retrieval relative to retrieved verbally (replicated by Mulligan & Hornstein, 2003), while re-enactment does not benefit free or cued recall (Helstrup, 2005; Kormi-Nouri et al., 1994; Norris & West, 1993; Steffens, 2007). In fact, while some studies show an increase in memory when participants enact actions at retrieval after verbal encoding (Koriat, Ben-Zur, & Nussbaum, 1990; Kormi-Nouri et al., 1994; Norris & West, 1993), even if participants only gesture rather than explicitly re-enact (Cook, Yip, & Goldin-Meadow, 2010; Stevanoni & Salmon, 2005), others do not (Brooks & Gardiner, 1994; Saltz & Dixon, 1982). This suggests that re-instatement of motor information at retrieval benefits identification of actions as performed or not-performed in recognition tasks, but does not benefit retrieval (i.e. accessing the items in memory) of those actions in recall tasks.

Engelkamp et al. (1994) suggested that the enactment benefit after re-enactment of motor action verbs was greater when the same hand was used to perform the action at both encoding and retrieval; Mulligan and Hornstein (2003) could not replicate this hand-specificity effect with verb-object phrases. It is possible that memory retrieval in Mulligan and Hornstein was driven largely by object identity rather than motor action, while the focus in Engelkamp et al. was necessarily more on the movement (since participants encoded verbs not verb-object items).

An alternate account to the multi-modal theory is one that focuses on the function motor activation at encoding may fulfill in boosting memory.[1] While the multi-modal account of enactment (Engelkamp, 2001) proposes that the retrieval of the motor trace is central, this account argues that encoding actions by enactment simply leads to better integration of the components of the item in addition to better processing of the components (Kormi-Nouri et al., 1994; Steffens, Jelenec, & Mecklenbräuker, 2009). Rather than the motor trace being important in isolation (as the multi-modal account suggests), this integration account proposes that enactment simply focuses participants on the task-relevant features of the item, a function that could also be fulfilled by other manipulations. Indeed, once actions are enacted, manipulations of levels-of-processing (R. L. Cohen, 1981; Zimmer & Engelkamp, 1999), anticipation of learning judgements (R. L. Cohen, 1983), generation (Nilsson & Cohen, 1988) and conceptual or motoric enrichment (Helstrup, 1987; Nilsson & Cohen, 1988; Nilsson, Nyberg, Nouri, & Rönnlund, 1995) do not further moderate the size of the enactment effect. This suggests that the increased item memory following enactment may not be due to encoding of a separable, additional memory trace but merely due to the increased processing of the semantic information by enactment. Given the variety of action stimuli, experimental designs and retrieval tasks used, it is difficult to collate the evidence and trace the patterns of results to pin down the crucial mechanisms. However, the experimental research and manipulations during the encoding of actions when assessing item memory for actions, provides a good baseline for looking at the effects of similar manipulations on source memory for actions.

---

[1] Other early accounts include R. L. Cohen (1981) proposing that the enactment benefit is a result of the automatic, optimal encoding of items after enactment, while verbal encoding relies on acquisition and rehearsal strategies and is therefore more prone to failure. Bäckman et al. (1986) argued for a dual-code account of strategic encoding of the verbal component and automatic encoding of the motor action and object. Under their account, the implicitly encoded components can trigger retrieval of the explicit memory (Nilsson & Bäckman, 1989).

1.2 Source memory for actions

The enactment effect is defined as superior memory retrieval of items after enactment relative to verbal encoding or encoding by observation or imagination. The studies I discussed above therefore focused on the number of items participants retrieved, with results typically reported as proportions of items recalled or proportions of items recognized. While the studies using recognition paradigms reported hit rates corrected by false alarm rates (e.g. Engelkamp & Dehn, 2000; Golly-Haring & Engelkamp, 2003; Mulligan & Hornstein, 2003), the studies using recall paradigms only reported proportion recalled without mentioning the items participants falsely recalled (i.e. intrusions of novel items). While a number of enactment studies used between-subject paradigms where separate participants enacted or, for example, observed actions at encoding, others manipulated enactment and observation within-participants (e.g. Golly-Haring & Engelkamp, 2003). At test participants were asked to simply retrieve items they had encoded, so again, the focus of the analysis was on the overall proportion of items participants correctly recalled at test. In other words, the enactment effect focuses on how item memory for actions changes with different manipulations.

From here-on-in, I will discuss experimental work that focuses not on participants' item memory, either as quantity or accuracy of recall or recognition, but primarily on their source memory for actions. Source memory for actions is not typically discussed within the enactment literature, and when it is, it is framed as an investigation of the benefits of enactment on source memory performance. Throughout this thesis, I will approach the topic of source memory for actions from a different angle. Given that people are typically able to not mistake actions they observed as ones they had performed themselves, under which circumstances do these errors occur and which mechanisms are responsible for this confusion of having performed or having observed an action. Rather than looking at whether participants simply remember that an event occurred, I will look at participants' source errors, that is the events they falsely remember having performed when they observed someone else perform the action, or the events they falsely remember having observed when they performed an action themselves.

There are two different theoretical approaches I will discuss in detail that may explain participants' ability or failure to remember whether they performed, observed or even imagined an action.

The first approach is born out of source monitoring research in general and reality monitoring research in particular (Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981). Here source memory for actions is assumed to be subject to the same mechanisms and evaluations as source memory for verbal and pictorial items and result in similar patterns of source performance. This account, much like the multi-modal account proposed to account for the enactment effect (Engelkamp, 2001), assumes that memory for an item contains multiple features or traces. These traces or features are evaluated at retrieval to infer the source of a retrieved items, with some features more diagnostic of some sources than other sources (Leynes & Kakadia, 2013). I will discuss how this account explains source confusion of performed and imagined and performed and observed actions.

The second account I will discuss is a novel, motor simulation account (Lindner et al., 2010), that has been proposed to account specifically for the false retrieval of observed actions as performed. It is based on neuroscientific evidence that suggests that not only performed, but also observed actions contain motor representations. According, source confusion here occurs because motoric encoding during observation leads to the creation of motor memory traces that indicate that observed actions were self-performed.

Throughout my discussion, I will touch on alternate memory accounts that propose that overall memory strength, rather than separable features, can account for source memory performance. Under the relative strength account (Hoffman, 1997; Marsh & Bower, 1993), participants' source judgement is based on items' overall memory strength, with the average memory strength higher for some sources. Here misattributions occur because stronger items from the weaker source are misattributed to the stronger source and vice versa.

1.2.1 The source monitoring account

Source monitoring is people's ability to monitor the source of information they retrieve from memory, by distinguishing between external sources (whether they read something in the newspaper or heard it on TV) or internal sources (whether they imagined or they dreamed something). According to the source monitoring framework (Johnson et al., 1993), the source of an item is not encoded explicitly alongside the item as a source tag. Instead, participants' memory representations of an item contain not only the semantic item information but also qualitative, episodic features of the encoding event, such as perceptual, temporal, spatial, cognitive, affective or sensory characteristics, i.e. thoughts participants had during the event, what an object felt like, the emotional state they were in or the context of the event. When it is necessary to determine the source of a remembered item, people infer the source on the basis of those qualitative features in a heuristic process since different sources are characterized by different qualitative features (Johnson & Raye, 1981). For example, imagined items are characterised by strong cognitive traces while items witnessed in the environment are rich in perceptual information (Hashtroudi et al., 1990; Suengas & Johnson, 1988). Additionally people can infer the source more systematically, if necessary, by questioning whether it is plausible that this item would arise from that source (Johnson, 1985). Source judgements are made based on a combination of these heuristic and systematic processes. Note that the dissociation of item memory (whether the item can be retrieved at all) and source memory (whether the source of the item can be identified) are explicit in this theoretical account. While memory for the item (semantic memory) and the retrieval of features may be correlated, they are assumed to be separate.

Failures in source attribution arise when qualitative features are not encoded or are not considered appropriately at retrieval. Source failures occur for example when encoding under divided attention (e.g. Jacoby, Woloshyn, & Kelley, 1989; Lane, 2006) or when the focus is only on a selection of features (Johnson, Nolde, & De Leonardis, 1996). Similarly, source misattributions increase when retrieval takes place under divided attention (e.g. Dodson, Holland, & Shimamura, 1998; Dodson & Johnson, 1996; Zaragoza & Lane, 1998), time pressure (Zaragoza & Lane, 1998), or when response options at test bias responding (e.g. Chan, Wilford, & Hughes, 2012; Dodson & Johnson, 1993; Marsh & Hicks, 1998).

Importantly, a prediction of the source monitoring framework is also that source failures increase with increasing similarity of sources. Johnson, Foley, and Leach (1988) showed that distinguishing imagined from heard words led to more source misattributions when participants were instructed to imagine the words in the external speaker's rather than their own voice. While the imagined versus heard voice cue was indicative of one or the other source when the speaker differed for sources, it was not diagnostic of source when the same speaker was imagined as heard. Similarly, hearing words spoken by participants of the same gender increases source confusion (Lindsay, Johnson, & Kwon, 1991). Durso and Johnson (1980) showed that instructions to imagine an object when shown the word increased participants' tendency to falsely identify items that had been presented as words with items presented as pictures. Similarly, instructions to elaborate on perceptual features of an event lead to increased identification of the event as experienced rather than presented post-hoc (Drivdahl & Zaragoza, 2001). Similarity of to-be-learned objects increases false memories for location (Lyle & Johnson, 2006), while similarity of to-be-learned items increases confusion over who participants heard say the item (Lindsay et al., 1991).

The source monitoring framework has been applied to a number of source memory phenomena such as source confusion in eyewitness statements (e.g. Zaragoza & Lane, 1998) and cryptomnesia. In the standard cryptomnesia (or unconscious plagiarism) experiment (based on Brown & Murphy, 1989), participants in groups take turns to generate solutions for a task. Following a delay participants are asked to complete a recall and/or a generate-new task. In the recall task, participants are asked to selectively recall the solutions they generated themselves, avoiding those generated by others in the group. In the generate-new task, participants are asked to generate novel solutions to the task, avoiding both previously self- and other-generated ones. Plagiarism errors (or source errors in the recall-own and generate-new task) are now solutions generated by other members of the group that participants falsely claim to have generated themselves, with plagiarism typically at above-chance rates for both the recall-own and generate-new task (e.g. Brown & Murphy, 1989). This unconscious plagiarism effect has been shown when participants generate exemplars for semantic or orthographic categories (e.g. Brown & Murphy, 1989; Hollins, Lange, Berry, & Dennis, 2016; Macrae, Bodenhausen, & Calvini, 1999), create anagrams (e.g. Marsh & Bower, 1993), create alternate uses for objects (e.g. Stark, Perfect, & Newstead, 2005; Stark & Perfect, 2006) or create solutions for real-world problems (Perfect, Field, & Jones, 2009). In line with predictions from the source monitoring framework (Johnson et al., 1993), unconscious plagiarism increases with delay (e.g. Hollins et al., 2016), distraction at encoding (Macrae et al., 1999), similarity of sources (Defeldre, 2005; Hollins, Lange, Dennis, & Longmore, 2015; Landau & Marsh, 1997; Macrae et al., 1999) and under speeded responding (Marsh & Bower, 1993). Unconscious plagiarism decreases with source cues at retrieval (e.g. Hollins et al., 2016; Marsh & Bower, 1993), with warning of the effect (Marsh & Bower, 1993) or instructions to consider the source of retrieved ideas (e.g. Marsh, Landau, & Hicks, 1997). More recently, Hollins et al. (2016) showed that source errors in a recall tasks are not limited to the recall-own task, but also occur in the recall-partner task. Rather than participants being biased to simply claim ideas as their own, it appears that participants are simply confused about the source of the ideas they retrieve from memory (see aso Hollins et al., 2015; Perfect et al., 2009). This research on unconscious plagiarism shows that a) source errors are not simply a result of the source recognition task at test and b) the source monitoring framework can be applied to source memory measured in recall paradigms.

While the source monitoring framework was not explicitly conceptualized to account for source confusion of verbal material, the majority of the experimental work I discussed thus far focused on the source confusion of words or verbalized ideas. In my thesis, I will look at the source confusion of actions. In the next section I will discuss the literature on source memory for actions, as it has been discussed within the source monitoring framework. I will briefly touch on source memory for performed and imagined actions (known as reality monitoring failures within the source monitoring literature (Johnson & Raye, 1981)), before I move on to discuss the existing research on source memory for performed and observed actions.

Confusing self-performed and imagined actions

The source monitoring framework predicts that high perceptual overlap between sources should increase source monitoring errors. This can account for why imagined actions are frequently mistaken as having been self-performed (Hashtroudi et al., 1990; Johnson & Raye, 1981; Lampinen, Odegard, & Bullington, 2003). For example, Garry, Manning, Loftus, and Sherman (1996) reported that when participants were told to imagine a fictional childhood event, they later claimed this event had occurred with higher confidence than fictional events they had not been instructed to imagine.

One central paradigm that was developed to investigate this effect of imagination on false memories of self-performance in the lab is the imagination inflation paradigm. Goff and Roediger (1998) asked participants to listen to action phrases like ‘Holepunch the paper’ in the initial encoding phase. Some of these actions, participants were asked to perform using objects, some they were asked to imagine performing and some they only heard. In a second encoding phase some of the actions presented in the initial encoding phase and novel actions were read to participants. Participants were asked to imagine themselves perform these actions. Some of the actions in this second encoding phase were presented once, some three times and some five times. Two weeks later, participants returned for a written recognition test. For each item, participants first decided if it had been presented in the initial encoding phase. For items they identified as originating in the initial encoding phase, they then answered a 3 alternative forced choice (Heard, Imagined, Performed) source recognition question.

Goff and Roediger (1998) focused their analysis on false claims of self-performance after imagination in the second encoding phase, when participants had heard or imagined the action in the initial encoding phase, or when it had not even been presented in the initial encoding phase. Increased number of imagination trials in the second encoding phase increased the rate of false ‘Performed’ responses for actions not presented in the initial encoding phase. The effect was not significant for actions that had been heard or imagined in the first encoding phase according to the authors, but was significant when the analysis was collapsed across all three types of non-performed actions (actions not presented at encoding, heard or imagined). The authors did not report if imagining actions also increased false ‘imagined’ responses at test.

While these findings are in line with predictions of the source monitoring framework, Goff and Roediger (1998) proposed an alternative account. They argued that asking participants to imagine performing actions may simply increase the memory strength of these items. Since actions performed in the initial encoding phase have higher memory strength (familiarity) than actions merely heard, this high memory strength of imagined actions would lead to false source attributions of those actions as performed (in line with a mistaken fluency account Jacoby, Woloshyn, & Kelley, 1989).

Thomas, Bulevich, and Loftus (2003) (Experiment 2) tested this memory strength account of the imagination inflation effect. In a paradigm based on the one used by Goff and Roediger (1998), participants were asked to perform or imagine actions in the initial encoding phase. In the second encoding phase, participants imagined themselves performing actions, with either simple imagining instructions as in Goff and Roediger or elaborate imagining instructions (focus on sensory details), or they merely read the action statements. Two weeks later, participants completed a two-stage recognition test, where they first determined whether an action had been encoded in the initial encoding phase, and then determined if they had performed or imagined that action in the initial encoding phase (for actions they identified as originating in the second encoding phase). They additionally asked participants to provide Remember/Know judgements for their source decisions. Participants were instructed to use ‘Remember’ for actions for which they had a vivid memory of the exact occurrence and ‘Know’ for actions they were certain about but could not recollect explicitly.

Both simple and elaborate imagination led to comparable rates of false ‘Performed’ responses, with ‘Remember’ responses frequently given. Merely reading action phrases in the second encoding phase increased false ‘Performed’ responses to a lesser degree. Based on this, Thomas et al. (2003) argued that the sensory elaboration of imagining actions was a necessary prerequisite for the inflation effect. In other words, enriching the memory trace with sensory features led to more source confusions (in line with a source monitoring account Johnson et al., 1993). According to the authors, mere familiarity with actions (under the mistaken fluency account; (Jacoby, Woloshyn, & Kelley, 1989)) could not account for it, or simply reading actions would have led to similar inflation effects. Subsequently, (Thomas & Bulevich, 2006) showed that when participants were asked to systematically evaluate the qualitative or sensory features of their memories of actions at test, the inflation effect reduced. Arguably, neither study provides conclusive evidence against a memory strength based argument of imagination inflation. The lower rate of false ‘Performed’ responses after reading action phrases may simply be because the memory strength for read actions is lower than for imagined actions.

If the featural overlap created between performed and non-performed actions by imagination led to false attributions (source monitoring account), how or who participants precisely imagined performing an action should affect the rate of false claims of self-performance. Seamon et al. (2006) showed that imagining an action led to false claims of performance, both for confusions of self-imagination with self-performance and other-imagination with other-performance (they did not test the cross over effect). Lindner and Echterhoff (2015) (Experiment 3) showed that other-imagination can lead to false memories of self-performance. Though there was a trend towards claims of self-performance being larger after self-imagination, this was not significant. For claims of other-performance, on the other hand, imagining someone else perform the action was more likely to make participants claim they had observed the action being performed, than if they had imagined themselves performing the action. This suggests that the preciseness of the featural overlap does impact the extent to which imagination leads to false claims of performance.

False memories of self-performance also occur when imagination is not explicitly instructed. Henkel (2011) showed that seeing photos of completed actions such as a stapled piece of paper for ‘Staple the paper’, increased false memories of self-performance, while Lindner and Henkel (2015) reported that simply hearing actions being performed, like paper being stapled, can lead to participants later claiming they had performed these actions. In both cases, some type of sensory or visual elaboration took place, relative to mere re-exposure to an item, in line with the argument made by Thomas et al. (2003).

Confusing self-performed and observed actions

There are only a few studies that looked at participants’ ability to distinguish self-performed from observed actions. The studies, and the reported rates of source errors, are shown in Table 1.2. I will discuss them individually below. This specific type of source confusion has been tested in two different paradigms: a standard source recognition paradigm and the observation inflation paradigm (adapted from the imagination inflation paradigm). I will discuss the standard source recognition studies first before discussing the novel theoretical motor simulation claims arising from studies using the observation inflation paradigm.

In the first three studies I discuss, participants’ source memory for performed and observed actions was measured, but relative rates of source errors were not compared.

Foley and Johnson (1985) (Experiment 1) tested 6 year olds, 9 year olds and adult participants. They asked participants to take turns self-performing and observing actions (the do-watch condition). All actions were simple, verbalizable body movements that were performed without objects, such as ‘Trace a B’ or ‘Look at the ceiling’. In addition to the do-watch condition, Foley and Johnson also tested different participants in the watch-watch condition (observing two different people perform actions) and the do-pretend condition (take turns self-performing and imagining actions). After a brief retention interval, all participants first completed a free recall test of the items that had been presented at encoding. Adults recalled more items than 9 year olds who recalled more items than 6 year olds, with no evidence for a difference in the number of items recalled between conditions (do-watch, watch-watch, watch-pretend). Foley and Johnson did not report recall quantity separately for the types of items in each condition in this experiment, nor did they assess participants’ recall accuracy by reporting false alarms or asking participants to selectively recall only items from one source. Instead they tested discrimination performance in a 3AFC source recognition test following the free recall test, in which discrimination was tested both for items participants had recalled and not recalled (with discrimination performance reported as equivalent and therefore collapsed across recalled and not recalled items).

In the do-watch condition participants had to determine whether actions presented at test had been performed by them, observed by them or were new. In the watch-watch condition participants were asked to distinguish who they had observed perform the actions, and accordingly in the do-pretend condition participants had to distinguish whether they had performed or imagined the action, or whether it was new. False alarms in this paradigm are novel actions falsely classified as old. False alarms occurred more frequently in the do-pretend condition than in other conditions, though Foley and Johnson (1985) did not list false alarms by source, so it is not possible to asses whether participants were biased in their Old/New recognition.

Discrimination performance was calculated as the average conditional source identification measure (average CSIM). The CSIM gives an index of source discrimination for items correctly identified as old. The average CSIM in the do-watch condition, for example, is then calculated as the number of self-performed actions correctly identified as self-performed plus the number of observed actions correctly identified as observed divided by the total number of actions correctly identified as old. Foley and Johnson (1985) did not report CSIM separately for performed and observed actions, so it is not possible to assess whether participants' source discrimination was driven by identification of one or the other source.

Adults showed comparable discrimination performance in all three conditions. In the condition critical for this thesis (the do-watch) condition, adults and children performed near ceiling (with an average source discrimination of .95 across all participants, or source discrimination failure of .05). Both 6 year old and 9 year old children only performed significantly worse than adults in the do-pretend condition.

Table 1.2. Source confusions in the control conditions of experiments as false claims of self-performance and false claims of observation reported in the literature

Study	Participants	Delay	Analysis	False ‘performed’	False ‘observed’
Source recognition					
Foley and Johnson (1985), † Experiment 1	Adults	immediate	average 1-CSIM	.09	
G. Cohen and Faulkner (1989) †	Adults	immediate	1-CSIM	.01	.03
Conway and Dewhurst (1995), † Experiment 2	Adults	immediate	Remember 1-CSIM	.02	.07
			Know 1-CSIM	.04	.1
Foley et al. (1993), Experiment 1	8yo	immediate	FSIM	.15	.09
Hornstein and Mulligan (2004)	Adults	1 week	1-CSIM	.13 (.11)	.12 (.08)
Manzi and Nigro (2008)	Adults	1 week	Remember 1-CSIM	.06 (.08)	.18 (.12)
			Know 1-CSIM	.03 (.04)	.14 (.07)
Rosa and Gutchess (2011)	Adults	immediate	FSIM	.04 *	.08 *
Leynes and Kakadia (2013)	Adults	immediate	FSIM	.03 *	.015 *
Observation inflation					
Lindner et al. (2010),† Experiment 1	Adults	2 weeks	Inflation	.13 *	
Lindner et al. (2012) †	Adults	2 weeks	Inflation	.16 *	
Schain et al. (2012) †	Adults	2 weeks	Inflation	.19 *	
Lindner et al. (2016),† Experiment 1	Adults	2 weeks	Inflation	.19 *	

Standard deviations in brackets where reported in the original paper

† - No formal comparison of rates of false responses to both sources

* - estimated from figures

1-CSIM - 1-conditional source identification measure; number of false source identifications divided by total number of items for a source correctly identified as old

FSIM - false source identification measure; number of false source identification divided by total number of items for a source

Remember - proportion of Remember judgements, Know - proportion of Know judgements

Inflation - difference of proportions of false ‘performed’ responses to observed and non-observed actions

G. Cohen and Faulkner (1989) (Experiment 1) asked adult participants (3 groups with mean ages 31, 65 and 76 respectively) to take turns performing, imagining and observing the experimenter perform actions. Objects were presented in an array of 24 objects at a time. Action phrases were created by manipulating two of the objects at a time, such as 'Put the toothbrush next to the cup'. Each object was used in only one action phrase. After a short retention interval, participants completed a 4AFC source recognition test (Performed, Observed, Imagined, New). Distractors were created by re-combining objects from the array into combinations that had not been used at encoding. Unlike Foley and Johnson (1985), G. Cohen and Faulkner reported rates of false alarms separately for 'Performed', 'Observed' and 'Imagined' responses. Participants were more likely to falsely identify distractor items as 'Observed' than as 'Performed' or 'Imagined'. This pattern of responding is referred to as the 'It had to be you' bias. Here novel items are misattributed more frequently to external than to internal events.

Bink, Marsh, and Hicks (1999) argued that this bias occurs because of participants' evaluation of features associated with events at retrieval, in line with the source monitoring account (Johnson et al., 1993). Novel items contain no features that are indicative of an internal event, but some features that are indicative of an external event. Hence, this featural misattribution leads to more false 'observed' than false 'performed' or 'imagined' responses for novel items. In contrast to this argument, Hoffman (1997) proposed that this bias is based on the relative memory strength of sources rather than evaluation of their features (also see Marsh & Bower, 1993). Under this account, novel items here are falsely attributed to have been 'observed' because observed actions result in weaker memory traces than performed or imagined actions. In other words, while novel items are strong enough to be falsely recognized as old, they are not strong enough to exceed the strength necessary to be falsely recognized as 'performed' or 'imagined'.

G. Cohen and Faulkner (1989) compared the CSIM between age groups and type of confusion separately for the three sources of old items. There was no evidence for participants being more likely to falsely identify performed actions as 'Observed' or 'Imagined'. For observed actions, only the oldest age group was more likely to misidentify observed actions as 'Performed' than observed actions as 'Imagined'. Misattributions for the two source were analysed separately, so G. Cohen and Faulkner did not report a formal comparison of false identification of observed actions as 'Performed' and performed actions as 'Observed'. Numerically, there does not appear to be a clear bias towards one type of false response over the other type of false response.

Conway and Dewhurst (1995) (Experiment 2) tested young adults using the same encoding paradigm as G. Cohen and Faulkner (1989) but employing a Remember-Know procedure at test. At test, participants were asked to give source judgements for items they identified as old. Following the source judgements, they were asked to assess the state of their memory for this item. If their source judgement was based on the recollection of details, they would respond 'Remember'. If it was based on a general feeling of knowing without recollection of details, they would respond 'Know'. If they had simply guessed they source, they were asked to respond 'Guess'. This method disentangles guessing from memory processes to an extent (at least as far as participants provide accurate metacognitive assessments of their memory). 'Remember' judgements are now closely associated with the recollection of features and details and thus map neatly onto the source monitoring account (Johnson et al., 1993). Source confusions contingent on item identification (CSIM) were analysed separately for Remember and Know judgements, but as in G. Cohen and Faulkner also separately for performed, observed and imagined actions. The authors found no evidence for different rates of source misattributions for Remember and Know responses overall. Numerically, false identification of observed actions as 'performed' was greater than false identification of performed actions as 'observed', but this comparison was not analysed statistically.

These studies suggest that participants both confuse performed actions to have been observed and observed actions to have been performed for actions that are performed with or without objects. While source performance appears high overall, this may be due to the short delay to test. Importantly, while the above studies looked at source confusion for the different sources of items, they did not explicitly compare whether source errors occurred more frequently in one than the other direction.

Foley et al. (1993) tested children's source memory (three groups with mean ages of 4, 6 and 8) for actions they performed and actions they observed an adult perform in a collaborative puzzle task. Each child participant took turns with an adult experiment to glue a total of 4 puzzle pieces each on a posterboard. Children participated in two conditions: the identifiable condition where they were asked to create a collage of a bear or bunny and the abstract condition where they were not given an instruction on what to create. After a brief retention interval, children completed a 2AFC (performed, observed) source recognition test. They were shown the collage they had made and asked to identify for each piece of the collage whether they or the adult had contributed that piece. Foley et al. analysed the frequency of source misattributions children committed. Overall 4 year olds committed more errors than 6 or 8 year olds, while performance between the older age groups did not differ. Across all age groups and both conditions (bar for the 6 year olds in the abstract condition), participants were more likely to claim they had performed actions they had observed than the reverse. This effect appears to be driven by the 4 year old participants, but Foley et al. do not report on any interactions. This suggests that children show an overall bias to claim actions they observed as self-performed.

Hornstein and Mulligan (2004) asked participants to take turns to perform actions and observe the experimenter perform actions. All actions were simple actions such as ‘Staple the paper’ using objects. They manipulated the amount of visual feedback participants received when they performed their actions. During self-performance some participants were instructed to close their eyes, some were instructed to perform actions with their eyes open and some were instructed to watch themselves perform the actions in a mirror. After two days, participants completed a 3AFC (Performed, Observed, Novel) source recognition test. Participants showed a clear enactment effect. They also showed the ‘It had to be you’ bias, where they were more likely to attribute novel actions to have been observed than self-performed. Visual feedback had no impact on participants’ item recognition, that is the correct identification of old actions as old. Hornstein and Mulligan analysed source discrimination using the CSIM (that is, correct source identification contingent on correct recognition of items as old).

Hornstein and Mulligan (2004) reported that participants both claimed to have performed actions they only observed and claimed to have observed actions they performed themselves. The errors were moderated by visual feedback manipulation. Increasing visual feedback decreased participants’ ability to discriminate performed and observed actions. While CSIMs for performed and observed actions did not differ significantly for the eyes closed and standard eyes open condition, they did differ when participants observed themselves in the mirror. Observing themselves in the mirror, decreased participants’ correct source identification of performed actions. At the same time, visual feedback during performance had no significant effect on participants’ source judgements of observed actions. Thus it was not the case that the decrease in correct ‘Performed’ responses after observation in the mirror was simply the result of an overall increase in ‘Observed’ responses (i.e., due to a bias in responding). Interestingly, it was not the case either that participants' source performance decreased overall due to an increased perceptual similarity of sources. Rather, in line with a source monitoring account that stresses the diagnosticity of features (Johnson et al., 1993; Leynes & Kakadia, 2013) the increased similarity of performed actions with observed actions selectively led participants to falsely identify more performed actions as observed. To reiterate, there was no evidence for a difference in the CSIM for performed and observed actions in the standard condition (though note that item recognition was higher for performed than observed actions, so the source discrimination measures are confounded by item recognition (e.g. Murnane & Bayen, 1996). Note that source memory performance but not item memory performance was affected by the experimental manipulation. This provides some evidence for the dissociation of item memory and source memory that the source monitoring account proposes.

Manzi and Nigro (2008) tested how delay would effect source memory performance for performed and observed actions. Participants took part in the experiment in pairs. Participants sat facing one another across the table and took turns performing and observing actions, though they did not strictly alternate. For each item, one of the participants was read an action phrase like ‘Staple the paper’, given the necessary object and instructed to perform the action. Participants returned for the retrieval test after 1 week or 2 weeks. Here participants completed an recognition test starting with an Old/New judgement for an item followed by 3AFC (Remember, Know, Guess) awareness of memory judgement and 2AFC (Self, Other) source test for items identified as old. Manzi and Nigro reported an enactment effect, that is higher Old/New recognition for performed than observed actions. They also reported that participants overall showed a bias to respond ‘Other’ for actions correctly identified as old (that is to claim they had observed their partner perform actions), for all memory of awareness levels (Remember, Know). This means that while participants were more likely to disown self-performed actions, they were also more likely to correctly identify observed actions as ‘Observed’ than performed actions as ‘Performed’, suggesting an overall bias to respond ‘Other’. Contrary to predictions of the source monitoring framework, delay did not decrease source performance (Johnson et al., 1993), though it is possible that a comparison of immediate test and a week's delay would have shown a different pattern of results.

In Rosa and Gutchess (2011) groups of three participants attended the encoding session together. Two participants knew one another while the third was unknown to both of them. The authors investigated whether source confusion for actions would be higher between close than unknown participants. They separately tested participants from two age groups (mean age 21 for younger and mean age 76 for the older participants). The triads completed two tasks: packing a suitcase and packing a picnic basket. Participants took turns placing items into the suitcase or picnic basket and observing their partners. After a brief delay, participants completed a 4AFC (Self, Close other, Unknown other, New) source recognition test. Rosa and Gutchess separately analysed correct source identification measures, false source identification measures and item recognition. There was no significant evidence for an enactment effect (better item recognition for self-performed than observed actions). Older adults committed more errors overall, but the closeness of the interaction partners had no significant effect on memory performance. Participants showed higher correct source identification for actions they performed themselves. The authors analysed incorrect source identification across all possible combinations of errors. Given the number of comparisons, they only reported that participants were less likely to misattribute the source for self-performed actions than observed actions, but this comparison does include participants confusing both their partners' contributions frequently (in line with predictions of the source monitoring framework (Johnson et al., 1993)). Examination of the figure in Rosa and Gutchess suggests that participants were more likely to falsely identify self-performed actions as 'Observed' than observed actions as 'self-performed'.

Finally, Leynes and Kakadia (2013) tested differences in participants' ability to distinguish performed and interrupted and performed and observed actions. Participants were presented with simple action phrases like 'Complete the puzzle'. For each trial, they were instructed to either perform the action (60 actions), observe the experimenter perform the action (30 actions), or perform the action and were interrupted mid-performance by the experimenter (30 actions). After a brief delay participants completed first a 3AFC (performed, observed, new) source recognition task followed by a 3AFC (performed, interrupted, new) source recognition task. The authors analysed item recognition by discrimination d' (contrasting normalized scores of hits and false alarm rates) and source performance by analysis of proportions of correct source identification, source missatributions and misses. They reported that participants were more likely to correctly identify performed actions as old than observed actions (the enactment effect). Leynes and

Kakadia analysed source performance across both source monitoring tests. Inspection of the figures reveals that participants were more likely to correctly response ‘Performed’ to performed actions (than ‘Observed’ to observed actions). If anything, there was a trend to falsely respond ‘performed’ to observed actions more frequently than ‘observed’ to performed actions, suggesting that source responding may be driven by a bias to claim actions as ‘performed’. In line with predictions from the source monitoring framework that predicts that increasing similarity of sources will lead to increases in source misattributions (Johnson et al., 1993), interrupted actions were more likely to be misattributed to have been performed than observed actions.

Leynes and Kakadia (2013) proposed a diagnostic features account of source memory for performed and observed actions to explain the source confusion between performed and observed actions. In their interpretation of the source monitoring account, both sources of actions are characterized by certain characteristics. While the presence of the motor trace, somatosensory trace and cognitive trace is diagnostic of an action having been performed, the presence of a third-person visual perspective trace would be indicative of an action having been observed. In line with predictions from the source monitoring framework (Hashtroudi et al., 1990; Johnson et al., 1993; Lampinen et al., 2003), actions would be more likely to be misattributed if participants did not encode certain characteristic features or if features typically characteristic of one source of actions would be encoded with the other source of actions. I will use this framework to guide the experimental exploration in my thesis since it allows direct testing of predictions about the relative importance of individual source features.

Additional studies investigating source memory for performed and observed actions studies have been conducted with cognitively impaired adults (Rosa, Deason, Budson, & Gutchess, 2014) and adults with schizophrenia (Brodeur, Pelletier, & Lepage, 2009). Beyond that, the studies listed here are, to my knowledge, the only studies that have looked at this particular kind of source confusion using the standard source recognition paradigm. The pattern of results (such as whether an enactment effect was observed, how frequent source errors were, whether false responding was biased towards one source) varies drastically between studies, due to the differences in method, actions used, retrieval task and analysis of data. However, two conclusions can be drawn from this research: (1) in all cases, participants misattribute performed actions to have been observed as well as misattribute observed actions to have been performed and (2) the effects of the manipulations largely follow predictions of the source monitoring framework (Johnson et al., 1993).

1.2.2 The motor simulation account

The source monitoring framework appears to be able to account for source confusion of action memories across a variety of paradigms. An argument for a more simple memory-strength-based explanation of source misattribution can be made given some of the data, and I will discuss that below. More recently an alternate theoretical account has been proposed to account for false memories of performance after observation that originates not in the memory literature but in the motor action literature.

Lindner et al. (2010)

Lindner et al. (2010) adapted the imagination inflation paradigm (Goff & Roediger, 1998) for observation. In the initial encoding phase, participants read simple action phrases like ‘Holepunch the paper’ on the computer screen. Participants were asked to perform half the actions using the necessary objects and only read the other half of actions. They asked participants to perform half of these action phrases and only read the other half. In the subsequent, second encoding phase, participants were again asked to read action phrases. Some of these action phrases participants had previously performed, some they had previously read and some were novel in this second encoding phase. In addition to reading the phrase, participants were asked to watch a video of someone performing the action using the necessary object. In the video the actor was shown from the neck down seated behind a table, with the video focused on the table top. The video showed the actor use the object to perform the action. The video looped repeatedly for a fixed video presentation of 15 seconds. Half the actions presented in this second encoding phase were only presented once, half were presented 5 times. Order of presentation was randomized.

After two weeks, participants completed a two-stage source recognition test that included items presented during encoding (both initial encoding and second encoding phase) and distractor items that were novel at test. For each item, participants first made a 2AFC (initial encoding phase, not initial encoding phase) judgement. For items participants identified as originating in the initial encoding phase, they additionally made a 2AFC (Performed, Read) judgement about how they had encoded the action phrase initially.

The authors were exclusively interested in the size of the inflation effect, that is, the increase in false ‘Performed’ responses resulting from observing someone perform actions in the second encoding phase relative to the baseline of false ‘Performed’ responses given to actions not observed in the second encoding phase (i.e., actions only presented in the initial encoding phase or actions only presented at test). In other words, does observation in the second encoding phase increase participants’ tendency to falsely identify actions as ‘Performed’ that they only read in the first encoding phase or that were not presented in the initial encoding phase. Lindner et al. (2010) reported a significant inflation effect after observation and named it the observation inflation effect (in analogy to the imagination inflation effect (Goff & Roediger, 1998)). Note, in this analysis, it is not clear if inflation is the result of participants’ failure to discriminate the sources of actions or simply reflects biased responding.

In addition to this observation condition, Lindner et al. (2010) also tested an imagination condition identical to Goff and Roediger (1998), a read condition and a generate condition with different sets of participants. In the read condition, participants merely read action phrases in the second encoding phase. In the generate condition participants were asked to unscramble the action phrase, i.e. unscramble ‘Sie3Schütteln9die5Flasche’ or ‘d__ fl_sche sch_tt_ln s_e’ to ‘Schütteln Sie die Flasche’ in the German original.

Lindner et al. (2010) observed a significant imagination inflation effect, replicating Goff and Roediger (1998). They also observed a significant observation inflation effect, which was not significantly smaller than the imagination inflation effect. Merely reading action phrases or generating them from a scrambled text did not increase false ‘Performed’ responses beyond the level observed at baseline, suggesting that mere re-exposure, even effortful re-exposure, is not driving the inflation effects of observation and imagination.

That observation leads to false memories of self-performance is still compatible with the source monitoring account. Performed and observed actions share a number of features, and may be confused easily. Lindner et al. (2010) subsequently tested a number of predictions from the source monitoring framework and failed to find significant evidence in its favour. In Experiment 2, they tested whether warning participants about the observation inflation effect would eradicate it. It did not. Neither did instructing participants to consider the perceptual features of the items eradicate observation inflation. Neither condition significantly differed from the standard instructions condition. In comparison, McDermott and Roediger (1998) showed that warning instructions can reduce the occurrence of false memories for verbal material and Thomas and Bulevich (2006) showed a reduction in the imagination inflation effect after instructing participants to focus on the perceptual features of items at test.

In Experiment 3, Lindner et al. (2010) tested whether increasing the perceptual similarity between performed and observed actions would increase source confusion (in line with predictions from a source monitoring account Johnson et al., 1993). Observing someone from a first-person perspective did not increase observation inflation relative to the standard condition, contrary to predictions. Neither did performing own actions with closed eyes significantly affect the size of the observation inflation effect.

In other words, the observation inflation effect remained impervious to source monitoring manipulations (i.e. resulted in null effects), leading the authors to suggest that false memories of performance after observation may not be the result of failures to infer the correct source of an item at retrieval based on the evaluation of qualitative features associated with that item. Lindner et al. (2010) suggested instead a motor simulation account of false memories of self-performance after observation due to mirror neuron activation in the observer's motor system.

Mirror neuron network activation

Mirror neurons were discovered first in macaque monkeys (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), with similar activation patterns subsequently observed in humans' motor system networks (Fabbri-Destro & Rizzolatti, 2008; Gallese, 2005). Mirror neurons are proposed to fire automatically (direct matching hypothesis; Iacoboni, 2005; Rizzolatti & Craighero, 2004) or in response to goals (action prediction hypothesis; Csibra, 2007) when an action is observed. Thus, observing an action engages some of the same neuronal populations as physically executing it (Bach, Bayliss, & Tipper, 2011; Bach, Peatfield, & Tipper, 2007; Brass, Bekkering, Wohlschläger, & Prinz, 2000; Oosterhof, Tipper, & Downing, 2013). The motor representations resulting from observation leave lasting traces of motor memory (Stefan et al., 2005) in the observer's motor cortex that match those created by action execution. This way observation creates internal replica of the action as if it had been self-performed (Grezes & Decety, 2001; Jeannerod, 2001).

By nature of the discussion, the majority of the mirror neuron network research is neuro-functional. The behavioural research that references mirror neuron network explanations focuses largely on concurrent or subsequent action performance and observation or imitation of small finger and arm movements (e.g., Brass et al., 2000; Kilner, Paulignan, & Blakemore, 2003; Press, Cook, Blakemore, & Kilner, 2011; Zwickel & Prinz, 2012). These experiments show that observing an action being performed facilitates own, compatible or congruent action execution (e.g. Brass et al., 2000; Zwickel & Prinz, 2012), and observing an action being performed changes the trajectory of own action execution in line with the observed trajectory (e.g. Kilner et al., 2003). While those experiments are designed to investigate the specific boundary conditions of the overlap of motor representations for performed and observed actions, the theoretical claims made by mirror neuron accounts extend into more complex predictions. For example, mirror neuron network activation has been proposed to form the neuronal basis for imitation learning, social cognition and empathy (Pfeifer, Iacoboni, Mazziotta, & Dapretto, 2008; Rizzolatti & Fabbri-Destro, 2008; Wohlschläger, Gattis, & Bekkering, 2003).

Lindner et al. (2010) argued that this account is compatible with the null effects of their experimental manipulations. Under their account of observation inflation, observing someone else in the experiment creates motor representations in the observer's motor system as described above (e.g. Iacoboni, 2005). They further elaborate: "Evidence suggests that mirrored motor representations can shape observers' self-related motor memory (Stefan et al., 2005) and that neural correlates of conscious memory for these representations are similar to those of memory for self-performed actions (Senkfor et al., 2002). When, on a memory test, participants reactivate mirrored action representations, they could – erroneously – remember having performed the action." (Lindner et al., 2010, p. 1297). In other words, they argue that the motor representation encoded in the observer's motor system reactivates at retrieval (as if the action had been self-performed) and leads to participants claiming an action as performed because the motor activation suggests it had been performed. Crucially, Lindner et al. (2010) do not argue that the motor trace encoded during observation simply increases the featural overlap between sources, meaning that source monitoring processes moderate the source attribution. Rather, they see "interpersonal motor simulation" (Lindner et al., 2010, p. 1297) as the critical process leading to errors in performance, with an instructed focus on perceptual features at retrieval (Experiment 2) or an increase in the overlap of features (Experiment 3) not moderating the inflation effect.

Lindner et al. (2010)'s interpretation of the mirror neuron network literature focuses on the interpersonal effects mirror neuron network activation has been claimed to contribute to, such as the development of social cognition and empathy (Pfeifer et al., 2008; Rizzolatti & Fabbri-Destro, 2008). Extrapolating from those interpretations, they argue that this explains why the first-person perspective in Experiment 3 (Lindner et al., 2010) did not increase observation inflation relative to the third-person perspective, but if anything showed a trend to decrease it. In their interpretation, the trend towards the lower inflation effect occurred "when the observation situation was less characteristic of social interaction" (Lindner et al., 2010, p. 1298). Note that the bulk of evidence in the motor action literature suggests the opposite, and shows greater impact of observed actions on the observer's motor system with first-person than third-person perspective observation (e.g. Bortoletto, Mattingley, & Cunnington, 2013; Watanabe, Higuchi, & Kikuchi, 2013).

Lindner et al. (2016) (Experiment 2) tested the prediction of the motor simulation account directly by limiting the motoric encoding of observed actions. They predicted that motor distraction at encoding, here: performing motorically incongruent actions during observation, would prevent encoding of the motor trace and thereby limit false performed responses at test. They asked participants to perform motor actions congruent or incongruent with the movement of the action they observed being performed. Observation inflation in the congruent condition was numerically higher (estimated at .25 from the table in Lindner et al.) than in the incongruent condition (estimated at .13). Rather than separately comparing only the inflation effects for false 'Performed' responses as in Lindner et al. (2010), the authors looked at the effect of inflation across correct and false 'Performed' responses. This means that it is not clear if the difference in observation inflation effect as originally defined is significant in this experiment. However, what the trend in the data suggests is that when participants pantomimed the action during observation, they were later more likely to falsely claim an observed action as having been performed by them in the first encoding phase. In other words, when participants performed the action in the second encoding phase, they were more likely to falsely claim they had performed the action in the initial encoding phase. Lindner et al. did not compare performance of incongruent actions during observation to a baseline no-performance condition. It is therefore not clear if incongruent actions disrupt motor simulation during observation and thus lead to a decrease in observation inflation, as the authors argued.

Further observation inflation studies

After arguing for the motor simulation account to explain their effects in Lindner et al. (2010), the authors reported results at least partially compatible with a source monitoring account. Lindner et al. (2012) reported that observation inflation decreased when participants observed an out-group rather than an in-group member (implemented by using actors with different skin colours) in line with research that showed decreases in motor simulation when observing out-group members (Gutsell & Inzlicht, 2010). Since inflation in neither in-group nor out-group condition significantly differed from the control perceptual-similarity condition where the actor in the video wore a glove, it is difficult to dismiss a perceptual similarity explanation for the effect (Johnson et al., 1993).

Schain et al. (2012) showed that seeing the face of the actor performing the action in the video decreased observation inflation (though did not eliminate it) relative to seeing the torso of the actor only, as in the standard paradigm. The inflation effect disappeared when participants did not pay attention to the action being performed but focused on the face only, though note that the conditions in which the face was visible did not differ significantly from one another. This suggests that providing non-self source cues to participants at encoding can reduce the rate of false performed responses. In other words, if participants encode cues that upon retrieval better indicate the source of an item, they are less likely to misattribute the source of that items, in line with the source monitoring account of source confusion (Johnson et al., 1993). Their results also show that actively encoding the action being observed is necessary for the observation inflation effect.

Finally, Lindner et al. (2016) (Experiment 1) tested the impact of the visual richness of the observed action performance on observation inflation. They proposed that the sensory overlap of features between self-performed actions and observed actions would be reduced when the video participants watched was impoverished visually. In line with the source monitoring framework (Johnson et al., 1993) they predicted a reduction in observation inflation. They presented the standard video of action performance in one condition and a strongly impoverished (high contrast) video in the other condition. They did not appear to observe a significant reduction in the observation inflation effect (estimated inflation of false ‘performed’ responses of .16 for standard videos and .12 from impoverished videos), though they did not compare this aspect of the data formally. This null-effect appears not in line with source monitoring framework predictions. However, arguably the opposite prediction would have also been compatible with a source monitoring account. Richer perceptual features are typical associated with external rather than internal events (Hashtroudi et al., 1990). In that case, the visual richness of the observed action should indicate that the action was observed more than the impoverished presentation, and the impoverished video should actually be more likely to lead to false ‘Performed’ responses - if the responding in the paradigm is due to uncertainty about correct source attributions. Rather than the experimental results not being compatible with the source monitoring account, the generic source monitoring account may simply be too flexible to allow for concrete predictions.

In sum some of the results presented using the observation inflation paradigm are compatible with the source monitoring framework, while others are not. Lindner et al. (2016) tested the central aspect of the alternative motor simulation account - the motor trace encoded during observation - and found no clear positive evidence in favour of this alternate account. Given the small number of studies that looked at source confusion for performed and observed actions at all, and therefore sparseness of evidence for the source monitoring account, this alternate account may present a viable alternative.

Bear in mind that one aspect that makes interpretation of the observation inflation data more difficult in comparison to the standard source recognition data is that participants perform a different source test. While participants in the source recognition task are asked to determine whether actions have been performed or observed, participants in the observation inflation paradigm are asked to distinguish performed or read actions (after those actions were observed or were not observed). While the measures typically used to analyse source recognition data confound source discrimination, item memory and guessing (e.g. Murnane & Bayen, 1996), the same is true even more so for data produced by the observation inflation paradigm. Given that Lindner et al. focus on rates of ‘Performed’ responses only, it is not clear if failures of discrimination or shifts in response bias drive performance in the experiments. While both could produce the results reported in those experiments, they would point to different underlying mechanisms. In the present work, I will explore some alternate ways to estimate memory performance in the tasks that control for guessing or biases.

1.3 Overview over the present work

The present literature on source memory for performed and observed actions is highly heterogenic with a variety of methods and retrieval paradigms as well as analytic approaches used to evaluate the data. This makes drawing clear conclusions about participants' ability to distinguish performed and observed actions difficult. In my thesis I will present a theory-guided, systematic test of theoretical accounts claiming to account for source confusions of performed and observed actions. In the next section I will first discuss my methodological approaches by presenting the retrieval tasks and analytic methods I will use in my experimental work. Following that I will summarise my experimental approach to testing predictions of the motor simulation account (Lindner et al., 2010) and a specific diagnostic features instantion of the the source monitoring account (Johnson et al., 1993; Leynes & Kakadia, 2013) of source memory for actions.

1.3.1 Methodological approach

Recall paradigms

In the literature so far, source memory for performed and observed actions has been explored using standard source recognition paradigms and observation inflation paradigms (see Table 1.2). To my knowledge, source memory for performed and observed actions has not been tested in the selective recall of actions. While recall studies have shown superior memory, that is, quantity of recall, for performed over observed actions in some experimental designs (the enactment effect; e.g. Golly-Haring & Engelkamp, 2003), studies in the action memory domain did not assess participants' ability to correctly recall actions from only one source without intruding with actions from the other source.

Standard selective recall task The standard selective recall task has been used extensively in the source memory literature with verbal stimuli, to explore source memory for self-generated ideas. As I referred to above, in unconscious plagiarism paradigms, participants take turns generating ideas, and are subsequently asked to only recall own ideas (i.e. selectively recall ideas from one source). They recall some of those ideas correctly, but also intrude with ideas their partner generated and novel items (for reviews see Gingerich & Sullivan, 2013; Perfect & Stark, 2008a). When participants are asked to selectively recall their partner's ideas they recall some of their partner's ideas correctly, but also falsely recall some self-generated ideas and novel ideas (Hollins et al., 2016, 2015; Perfect et al., 2009). I will adapt this unconscious plagiarism paradigm to use for actions to test if source confusion of performed and observed actions consistently extends to selective recall paradigms, and if the source confusion of verbal material that has been reported with this paradigm extends to actions.

The main concern of looking at source errors in recall tasks is identifying the genesis of these errors. While source errors in recognition tasks appear to follow directly from failures in discrimination or shifts in bias (though as Murnane and Bayen (1996) pointed out, those measures are also contingent on item recognition and guessing processes), source errors in recall tasks are more difficult to interpret.

The first issue is that source errors may result from guessing. This is a problem in particular when participants complete a forced recall task. In such a task, participants are asked to recall all items they encoded originally, and to guess if necessary. In a free recall task, on the other hand, participants are asked to recall as many items as they can and to terminate their recall when they cannot recall any more items. Tenpenny, Keriazakos, Lew, and Phelan (1998) showed that source error rates with verbal material increase with a forced recall task over a free recall task at test. Therefore, to limit errors occurring through guessing in my experiments, I will ask participants to perform a selective free rather than forced recall task.

In the interpretation of source errors, different techniques have been used distinguish source errors as a result of false memories and those arising from other processes. Rather than looking at absolute numbers of source errors, rates of source errors given the output at recall or rates of source errors in comparison to rates of intrusions of novel ideas are analysed (for a discussion see Perfect & Stark, 2008a). Further, there have been attempts to compare participants' production of source errors to chance performance. Chance performance has been estimated as the rate of self-repetition during generation or rate of participants' repetitions of generated ideas if they did not hear ideas being generated (Brown & Murphy, 1989). I will pursue a related approach in my work and estimate the number of source errors as a result of guessing by taking into account the typicality of items and the number of novel intrusions participants produce at retrieval. I will discuss my approach to this below.

A second issue that makes interpretation of source errors in recall tasks more difficult than in source recognition tasks is that errors in recall tasks are overtly the result of items

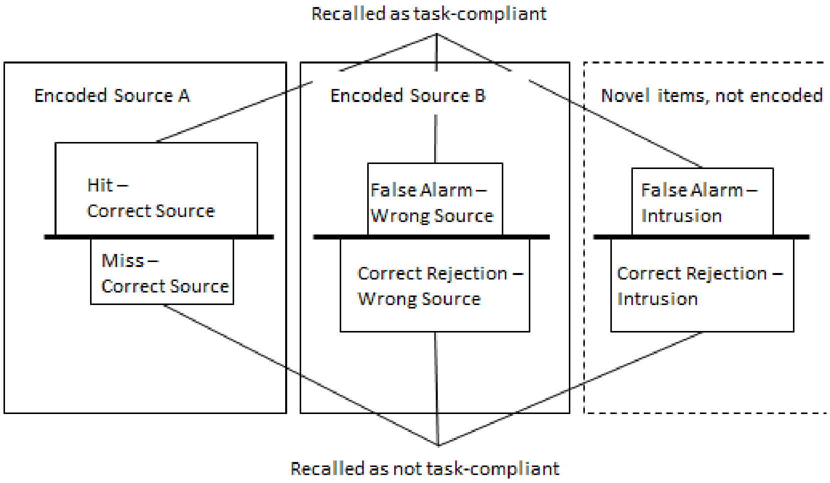

Figure 1.1. Responses in extended recall task. Responses in the top half are recalled as task-compliant (reported), responses in the bottom half are recalled as not task-compliant (produced at retrieval but withheld from report).

being available for recall (i.e., given the source cue participants are able to recall items from memory) and participants' ability to only selectively report items if they comply with the retrieval task. Source errors observed in recall tasks could be the result of either or both processes. In a source recognition task, on the other hand, participants ability to recall an item on cue from memory is never tested since participants' are presented with all items at test. In addition to the standard selective source-cued recall task, I will therefore also use an extended source-cued recall task.

Extended source-cued recall The extended recall task was first used by Bousfield and Rosner (1970) and more recently employed by Kahana, Dolan, Sauder, and Wingfield (2005) for a source-cued recall task. In Kahana et al., participants encoded word lists and were asked to recall only items from the last word list at test. They were given instructions to list all items that came to mind as they attempted to recall items from the last word list, and to indicate with a button press whether the item they just listed was an intrusion, i.e., an item not from the target word list. Hollins et al. (2016) adapted this extended source-cued recall paradigm to use in unconscious plagiarism. Participants took turns generating ideas at encoding. At retrieval participants were given a response sheet for a written recall task that contained a table with two columns. Participants were told that they were supposed to recall self-generated or were told they were to recall partner-generated ideas. However, they were instructed to write out all ideas that came to mind as they attempted to recall self-generated or partner-generated ideas. Whenever they recalled an item that complied with their retrieval task (for example, recall self-generated items), they were asked to write the item into one column, otherwise to write it in the other column. Figure 1.1 illustrates the possible responses in this recall task. At retrieval, participants may produce items from Source A, items from Source B or novel items. Some of these ideas they will report (top half in Figure 1.1). If Source A is the target source, items from Source A that participants report are correct responses, items from Source B are wrong source responses (source errors) and novel items are intrusion errors. These items would be written in one column. Items in the other column (bottom half of Figure 1.1) are the items participants withhold (or recall as not task-compliant). These may be items from their target source that would have been correct if reported, or items from the non-target source or novel items that are withheld correctly.

Hollins et al. (2016) analysed three different aspects of the data: the final report, the overall availability of items, and the monitoring performance. For the final report only those items participants reported as task-compliant were analysed. This is equivalent to the data from a standard source-cued recall task (though arguably with more explicit monitoring instructions). The overall availability of the items from a source was calculated by adding the number of items participants retrieved across both columns of the sheet, i.e. regardless of whether they identified the item as task-compliant or not. Monitoring performance was estimated by looking at the rates with which participants reported items as task-compliant (i.e., the number of items reported as task-compliant divided by the total number of items written down anywhere on the response sheet). Hollins et al. showed that self-generated items were more available than partner-generated items (in line with the generation effect (Mulligan, 2004; Slamecka & Graf, 1978)), and that source errors are influenced by both the availability of items at retrieval and the monitoring of sources.

I will adapt this extended source-cued recall for extended source-cued recall of actions to separately look at the effects of experimental manipulations on the availability of actions at test and participants' ability to monitor the sources of their memories. Since the availability of items largely maps onto memory strength and monitoring maps onto participants' ability to monitor the accuracy of responding, a dissociation of responses between these two stages of the retrieval task would provide some evidence for models assuming that item and source memory dissociate, such as the source monitoring framework (Johnson et al., 1993).

Guessing

One concern about source errors in any memory retrieval task is that source errors are a genuine memory error (the measure of interest) or simply a guess – an ad-hoc solution generated during the retrieval task – that just happened to be an item also generated at encoding. Source errors (i.e., false source responses to items correctly recognized as old) are therefore often analysed in relation to false source responses to novel items (i.e., false source responses to items falsely recognized as old) by measuring either the difference or the ratio of both types of errors. In fact, Lindner et al. (2010) analysed false performed responses following observation in relation to a baseline of false 'Performed' responses to items that were not observed. In their metric, this baseline represents guessing.

In Chapter 2, I will use a recall paradigm instead of the inflation paradigm (Lindner et al., 2010). This means I cannot use the same metric to account for accidental guesses. Instead, I developed a critical measure that would similarly take accidental guessing into account and look at an effect of observation beyond that, in a conceptual replication of Lindner et al. (2010)'s metric.

In my paradigm, I asked participants to generate actions, so I know which actions a lot of participants generate and which actions only one or two participants generate at encoding. This gives me a measure of fluency or typicality of actions. I used estimates of this typicality to model with a monte carlo simulation how frequently participants would also simply generate these actions at retrieval, i.e. if these actions are so fluent, how often would they come to mind as guesses? To estimate the effect of my manipulation I then contrasted the frequency of source errors I observed with the frequency of source errors I would expect if participants were only guessing. While this differs mathematically from the calculation of Lindner et al. (2010)'s guessing correction, it fulfills the same purpose.

Threshold and signal detection models

The majority of the experiments in my thesis are recall experiments. I outlined the analytical approach for these experiments above. In addition to the recall experiments, I will also present data from a standard source recognition and an observation inflation experiment. Source recognition data is typically analysed by calculating either the rates of responses given a source (SIM), or source discrimination between two sources given correct identification of items as old (CSIM). Murnane and Bayen (1996) pointed out that both measures confound item recognition, source discrimination and guessing (see also Batchelder & Riefer, 1990), with CSIM providing an unbiased estimate of source discrimination only if item recognition for both sources is equal. This tends not to be the case for performed and observed actions (the enactment effect; e.g. Engelkamp et al., 1994). This limits interpretation of the data in terms of isolating participants' source memory performance and comparing it across conditions. Murnane and Bayen suggested that source recognition data should be better analysed using multinomial processing tree models, with a two-high threshold model providing a good fit for source recognition data (e.g. Batchelder & Riefer, 1990; Bayen, Murnane, & Erdfelder, 1996; Murnane & Bayen, 1996). In this model, participants' responses are fitted by estimating item parameters, source parameters and guessing parameters.

Multinomial processing models make specific assumptions about the decision processes and memory representations underlying memory performance. These models distinguish discrete certainty and uncertainty states regarding item detection (whether an item is old or new) and source discrimination (whether an item belongs to source A or source B). Uncertainty states, i.e. if an item is not clearly identified, result in informed guessing. In multinomial models, probabilities of latent variables (that is, cognitive states) map onto the probabilities of observed responding at test. The assumption of discrete states results in rectangular memory distributions (e.g. Batchelder & Riefer, 1990; Bayen et al., 1996; Riefer & Batchelder, 1988).

Recently, there has been debate in the literature whether those discrete state assumptions accurately represent memory distributions. The alternative model tested against the Two-High-Threshold model in source recognition experiments is the signal detection model. Signal detection models assume that memory distributions are better represented by Gaussian distributions on a dimension of increasing memory strength (e.g. DeCarlo, 2003; Hautus, Macmillan, & Rotello, 2008; Jang, Wixted, & Huber, 2011).

Source recognition data can be accommodated by both threshold and signal detection models (e.g. Schütz & Bröder, 2011), so the debate centres on the curvature of ROC curves and slopes of zROCs for confidence rating data with models making different predictions given their assumptions about the distributions underlying performance (e.g. Batchelder & Alexander, 2013; Bröder & Schütz, 2009; Pazzaglia, Dube, & Rotello, 2013; Slotnick & Dodson, 2005; Yonelinas & Parks, 2007).

In addition to reporting hit rates, false alarm rates and source discrimination as typical with the source recognition paradigm, I will therefore also report model-based analyses by fitting two-high-threshold models (2HTM) and signal detection models to the data where appropriate. Since the model-free analysis does not provide a pure measure of source memory performance, comparing parameters estimates across conditions within either class of model will allow for a less biased interpretation. Rather than compare zROC slopes and ROC curvature to determine which model better accounts for confidence rating data, I will compare how well the models can account for response selection data and discuss model fits and parameter estimates (following DeCarlo, 2003; Schütz & Bröder, 2011).

While that aspect of the model-based analyses is motivated by the limits of the model-free analysis of source recognition data (that is analysis of the raw data), a second aspect of the model-based analyses I will present is based on the theoretical accounts I presented earlier. Under a source monitoring account, participants make source judgements by evaluating the qualitative features associated with the source. These judgements are separate from participants' memory for the item itself. An appropriate mathematical representation of the source monitoring framework is therefore a two-dimensional signal detection model where memory distributions are bivariate on separate item and source memory dimensions (e.g. Banks, 2000; DeCarlo, 2003; Hautus et al., 2008; Slotnick & Dodson, 2005; Wickens, 2011).

However, there are alternate accounts that could explain responding in source memory tasks. Participants may in fact not evaluate the qualitative features but base their responding on the overall strength of their memory representation. Under these theoretical assumptions, responding at test could be the result of participants simply producing the items they have the strongest memory for given a task (the fluency account; Jacoby, Woloshyn, & Kelley, 1989) or by comparing the relative strength of items from different sources (Marsh & Bower, 1993). These models are better represented by one-dimensional signal detection models. These one-dimensional signal detection models assume that responding is based on a single memory dimension and are in line with recent research that favours more parsimonious models to account for memory data (e.g. Berry, Shanks, & Henson, 2008; Shanks & Berry, 2012).

Throughout this thesis, I will fit both two-dimensional and one-dimensional signal detec-

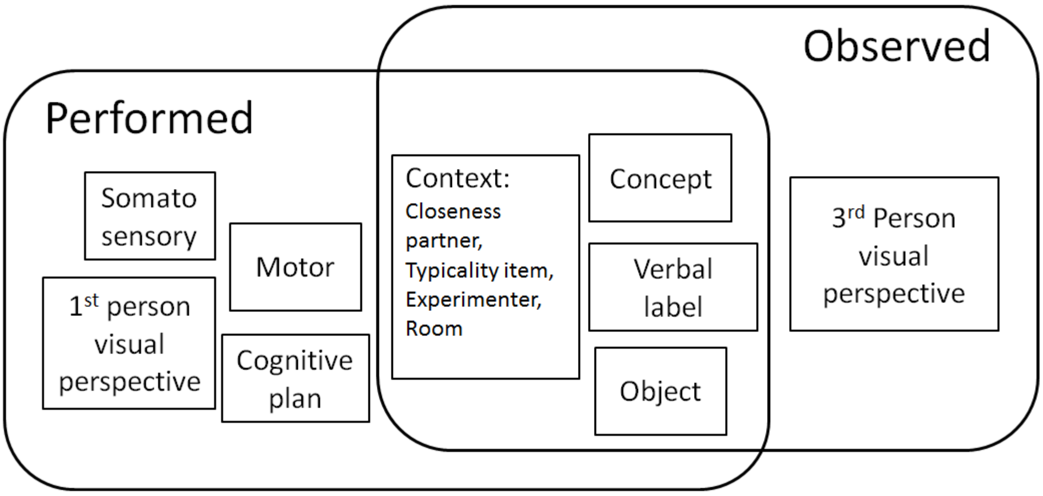

Figure 1.2. Qualitative features characteristic of performed and observed actions; adapted from Leynes and Kakadia (2013)

tion models to the data, to test whether the source monitoring or relative strength account better explain the data.

1.3.2 Experimental work

In the present work, I will explore how people are able to distinguish performed from observed actions and test some of the predictions that motor simulation and source monitoring accounts make. I will use the framework of performed and observed actions proposed by Leynes and Kakadia (2013).

Following a diagnostic features instantiation of the source monitoring account, Leynes and Kakadia (2013) propose that performed and observed actions are characterized by distinct qualitative features. Figure 1.2 shows an illustration adapted from Leynes and Kakadia. Under this account, performed and observed actions overlap considerably. Shared features include the verbal label and concept of the action, the object (if an action is performed with an object) and the encoding context. But there are features that are more diagnostic of performed actions and features that are more diagnostic of observed actions. Performed actions are characterized by somatosensory feedback, motor activation, cognitive activation and a first-person visual perspective. These are features that are not typically shared by observed actions. Observed actions, on the other hand, are typically characterized by a third-person visual perspective.

I will use this schema throughout this work to test if these qualitative features that distinguish performed from observed actions in theory, impact participants' ability to discriminate performed and observed actions. If overlap between sources increases, participants should be less able to discriminate these sources. This framework also predicts that specific features are diagnostic of specific sources. The presence of those features in items should then bias source responding to the source that those features are diagnotic for.

In Chapter 2 and Chapter 3, I will test the impact of the motor trace encoded during observation according to the motor simulation account proposed by Lindner et al.. Under this account, observing someone else perform an action leads to activation in the observer's motor system, resulting in memories enriched with a motor trace that will lead participants to believe they performed an observed action themselves. As Lindner et al. (2016) proposed, disruption of the encoding or retrieval of this motor trace for observed actions should result in a reduction of false memories of self-performance. To test the motor simulation account, I will first use motor actions that contain as little non-motor information as possible before moving on to more ecologically valid action items.

In Chapter 4, I will test the impact of the cognitive trace on source memory for performed and observed actions. Does increasing participants' cognitive involvement during the performance of actions increase their ability to distinguish performed and observed actions, as a source monitoring account would suggest? If the cognitive trace is diagnostic of performed actions, what is the impact of adding a cognitive trace to observed actions?

Finally, in Chapter 5, I will look at the effect of visual perspective on source memory for actions. This memory feature should allow clear distinction of performed from observed actions. Using the observation inflation paradigm, I will test if manipulation of the visual perspective at observation impacts participants' tendency to claim observed actions as self-performed.

Chapter 2

Motor component - encoding manipulation

2.1 Background

Source memory research typically uses the source monitoring framework (Johnson et al., 1993) to account for source confusion of verbal material. Lindner et al. (2010) proposed that a separate, motor simulation account can better explain false memories of self-performance after observation.

Using the observation inflation paradigm, Lindner et al. (2010) showed that participants falsely classify actions phrases they only read or that are novel at test as ‘performed’ if they observed someone else perform those actions. That finding itself could be accounted for within the source monitoring framework by arguing that items that were observed are more similar to performed actions than read or novel actions (Johnson et al., 1993). However, critically, in Lindner et al. (2010) the observation inflation effect did not decrease when participants were asked to consider the qualitative features associated with performed actions at test or when they were warned about the observation inflation effect (Exp. 2, Lindner et al., 2010). Additionally, increasing the perceptual overlap between performance and observation by manipulation of eyes closed/open and observing in 1st and 3rd person perspective did not significantly change the extent of the observation inflation effect (Exp. 3, Lindner et al., 2010).

This suggests that while the observation inflation effect itself is fairly robust, it appears impervious to manipulations of perceptual similarity and monitoring instruction. Both of these manipulations have been shown to affect source confusion of verbal material (Ferguson, Hashtroudi, & Johnson, 1992; Johnson et al., 1988; Thomas & Bulevich, 2006).

Lindner et al. (2010) argued that the crucial mechanism for this response shift is mirror neuron activation in the observer’s motor system during observation of action performance. Mirror neurons are proposed to fire when an action is observed (Csibra, 2007; Iacoboni, 2005; Rizzolatti & Craighero, 2004). The motor representations resulting from observation leave lasting traces of motor memory (Stefan et al., 2005) in the observer's motor cortex that match those created by action execution. This way observation creates internal replica of the action as if it had been self-performed (Grezes & Decety, 2001; Jeannerod, 2001).

According to Lindner et al. (2010), participants create an internal replica of an action as self-performed when they watch a video of an action being performed in second encoding phase of the observation inflation paradigm. At test, participants reactivate the motor information for the encoded actions, and activated motor information for merely observed actions leads them to falsely judge they had performed an action.

Lindner et al. (2010) propose that this motor simulation account can best explain their null effect findings following source monitoring manipulations. While this absence of evidence is hardly evidence of absence, manipulations that are theoretically predicted to affect source confusion in memory generally result in small effects in source memory research with actions. Hornstein and Mulligan (2004) asked participants to perform actions with their eyes closed, with their eyes open and when watching themselves in a mirror. They observed a linear trend of participants being more likely to confuse performed and observed actions with increased similarity of performed and observed actions, but individual comparisons of the conditions were not significant. While Rosa and Gutchess (2011) could not show a significant effect of source confusion increasing when observing a close versus an unknown other, Lindner et al. (2012) showed that source confusion increased when observing in-group versus out-group members, but notably, the effect in neither group differed from the control group matched for visual similarity.

While the trends in the data follow source monitoring predictions, they are not statistically or methodologically convincing. It is possible that a motor trace encoded during observation may particularly predispose observed actions to be mistaken for performed actions, beyond manipulations testing source monitoring framework predictions.

The critical component in the motor simulation account is the motoric encoding of observed actions in the observer's motor system. Lindner et al. (2010) tested predictions of the source monitoring framework (with the exception of Lindner et al., 2016, discussed later) and argued for the alternate motor simulation account when not finding evidence in favour of the source monitoring account. I will test the core component of the motor simulation account directly. If the motor component encoded during observation is critical for false memories of self-performance, preventing participants from encoding the observed actions motorically should prevent false memories of self-performance.

2.2 Methodology

In this first series of experiments, I developed a new paradigm to a) test whether the results reported by Lindner et al. (2010) generalize and b) test the motor simulation account of source confusion of actions directly. The paradigm I used in this series of experiments differs from the observation inflation paradigm in action stimuli, experimental design and retrieval task. If observation leads to false memories of self-performance, this should be the case independent of the experimental implementation.

Action stimuli

Lindner et al. (2010) used actions that are verbalisable as action phrases, as is common in action memory research. Unfortunately, verbalisable action phrases do not allow clear inference of the mechanism that leads to observation inflation: encoding a motor trace during action observation or enhanced verbal processing. For example, the verbal labels for observed actions may be remembered better because the action phrases could be clearly linked to external events, while those that were only read could not. At test, the only external event that was given as a response option was to say ‘Performed’. The majority of behavioural motor activation research, on the other hand, focuses largely on concurrent or subsequent action performance and observation or imitation of small finger and arm movements (e.g., Brass et al., 2000; Kilner et al., 2003; Press et al., 2011; Zwickel & Prinz, 2012).

To enable testing the role of the motor trace in creation of false memories of self-performance, I modified the action stimuli to be more in line with motor action research. The actions I used limit verbal encoding and leave only motor and visual encoding of the action. I asked participants to use any part of their body or combination of body parts to take turns performing actions in response to shape cues. This minimizes alternative non-motoric encoding strategies such as merely memorizing the objects used, and using them as cues to the actions associated with them. While action memory research has largely focused on enactment of action phrases as in Lindner et al., there are precedents for investigating memory of body movements such as dance moves and movement patterns (Foley et al., 1991; Helstrup, 1987, 1999, 2005; Smyth et al., 1988). Even though these actions are non-object-directed, this should not affect potential mirror neuron activation they invoke.

In fact, in humans motor system activation during action observation is typically higher in such non-object directed actions (Chong, Cunnington, Williams, Kanwisher, & Mattingley, 2008; Oosterhof, Wiggett, Diedrichsen, Tipper, & Downing, 2010; Press, Bird, Walsh, & Heyes, 2008).

Experimental paradigm and retrieval task

There are three major differences between the inflation paradigm and this novel paradigm: the action stimuli, the encoding task and the retrieval task.

Participants will perform and observe non-verbalizable actions intransient actions. In line with source memory for actions research (Hornstein & Mulligan, 2004; Leynes & Kakadia, 2013; Rosa & Gutchess, 2011), participants will take turns performing and observing actions at encoding. In contrast, in Lindner et al., actions are first read or performed, and only in a second phase additionally observed.[1] Observation there constitutes only additional encoding of already encoded action concepts. In the paradigm I propose, observation constitutes the only encoding of non-performed actions. In other words, the action is novel for participants at the time of observation. While participants in Lindner et al. watch videos of actions being performed, I will ask participants to observe actions in one-to-one interactions, as in, e.g., Hornstein and Mulligan (2004).

While actions in source memory for actions paradigms are encoded by enactment, they are typically tested verbally in a source recognition test. I wanted to know if the false retrieval of observed actions as self-performed generalizes beyond the observation inflation paradigm in particular and recognition paradigms in general to a free recall task. Lindner et al. (2010) used a two-stage source recognition task where participants first identify actions as old or new, and then decide whether actions they identified as old were performed or read. Later studies from the same lab employed a 3AFC (performed, read, novel) source recognition task where participants decided in a single-stage source recognition task whether items had been performed, read or not presented at encoding (Lindner & Davidson, 2014; Lindner & Echterhoff, 2015; Lindner & Henkel, 2015; Lindner et al., 2016, 2012; Schain et al., 2012). This is in line with source memory for actions research that asks participants to determine whether actions had been performed, observed or novel in 3AFC source recognition paradigms (Hornstein & Mulligan, 2004; Leynes & Kakadia,

---

[1] In Lindner et al. (2010), participants observed previously read, previously performed and novel actions. In subsequent research using the same paradigm, Lindner et al. did not present novel actions for observation. Participants only observed previously encoded (performed or read) actions.

2013; Rosa & Gutchess, 2011). Those studies establish that participants misattribute observed actions as performed in recognition paradigms, that is, when presented with an action they observed at test they falsely classify this action as self-performed. Will participants also misattribute the source of actions if they recall actions?

Similar generalizations of the enactment effect (in particular the advantage of self-performed over observed actions) proved less universal than initially assumed (Roediger III & Zaromb, 2010). While there are enactment effects in free recall tasks, they are more vulnerable to experimental design choices than enactment effects in recognition paradigms (Golly-Haring & Engelkamp, 2003; Steffens et al., 2015). It is possible that the observation inflation effect equally is driven by experimental design effects rather than underlying motor activation.

To my knowledge, there is no published research investigating the source confusion of actions in a free recall task but there are some studies on source confusion of verbal material. Here, the unconcious plagiarism effect or cryptomnesia (Macrae et al., 1999) describes participants falsely including partner-generated ideas when asked to recall self-generated ideas. In the prototypical paradigm (Brown & Murphy, 1989) participants take turns to generate solutions to verbal fluency or puzzle tasks. In a subsequent free recall-own task, participants are asked to recall the solutions they generated without incorporating any novel solutions or solutions their partners generated. Partner-generated solutions that are incorporated during this free recall-own task are called unconscious plagiarism errors or more generally referred to as wrong-source errors or simply source errors.

Rather than exploring the source confusion of performed and observed actions in a recognition task, I therefore decided to create a new recall paradigm to look at source confusion of action memories in a paradigm based on the unconscious plagiarism paradigm. Participants were asked here to take turns to perform and observe actions and at a later date recall some of the actions they performed themselves.

A consequence of using non-verbalizable actions in combination with a recall task is that I will ask participants to re-enact rather than verbally retrieve the actions they or their partner performed at encoding. Re-enactment at test results in small benefits for identifying performed and observed actions as old compared to verbal test (Engelkamp et al., 1994; Mulligan & Hornstein, 2003). It is possible that re-instatement of the motor information at test may enhance source memory performance relative to verbal retrieval, but there is no published research on the effects of re-enactment on source memory.

In addition to asking some participants to recall performed actions, I will also ask other participants to recall actions they observed their partner perform (in analogy to testing the recall of own and partner-generated verbal items (Hollins et al., 2016)).

2.3 Aims and predictions

The first question I want to address is if the observation inflation effect conceptually replicates in a recall task, even though actions cannot be easily verbalized and observed actions are solely encoded by observation. If observing actions generally results in false memories of self-performance, I would, as the critical measure, expect participants to falsely recall observed actions as self-performed, just as occurs in the verbal domain.

A second question concerns the direction of source misattributions. I know that participants not only report their partner's ideas as their own in the verbal domain (for reviews see Gingerich & Sullivan, 2013; Perfect & Stark, 2008a), they also commit the reverse error when given the chance and falsely recall their own ideas as their partner's (Hollins et al., 2016, 2015). If source memory for actions follows the same pattern, I would expect participants to commit source errors not only when they recall self-performed actions, but to also commit them when they recall actions they observed their partner perform. In other words, in addition to observation leading to false memories of self-performance, I expect that self-performance would also lead to false memories of observation. In fact, the source recognition studies for actions not using a inflation paradigm consistently show both errors (Hornstein & Mulligan, 2004; Leynes & Kakadia, 2013; Rosa & Gutchess, 2011).

Under a motor activation account during observation, only the observation inflation error (misremembering observed actions as self-performed) is easy to explain. Action mirroring creates a motor trace of the observed action that is added to its visual memory representation. During retrieval, there is conflict between visual and motoric memory traces, one suggesting observation and the other self-performance, which causes some of the actions to be misattributed. However, such views would be hard-pressed to account for observation of the reverse error, where participants misattribute an action to their partner that they had performed themselves. For self-performed actions, both motoric- and visual-memory indicate self-performance; there should thus never be a conflict about the source of a self-performed action.

This predicts a striking asymmetry: while people should readily claim others' actions as their own, they should very rarely do the reverse. This will be the first direct test of motor activation account predictions.

A final purpose of Experiment 1 is to provide a methodology for subsequently investigating the role of the motor trace during the observation of actions directly using a dual-task approach in Experiments 2 and 3. Under a motor activation view, I expect that interfering with the motor encoding of observed actions during observation should make those actions less likely to be confused for self-performed actions later.

2.4 Experiment 1

Experiment 1 aims to conceptually replicate the observation inflation effect reported by Lindner et al. (2010) in a free recall paradigm and with non-verbalisable actions. Additionally, I will test a key prediction of the motor activation during observation account that predicts that observed actions should be falsely retrieved as self-performed but self-performed actions not, or to a lesser degree, falsely retrieved as observed.

2.4.1 Method

Participants

37 members of the public participated for payment of £8. Two participants were excluded from analysis for not attending all sessions. The sample size was based on experiments in Hollins et al. (2016).

Procedure

Participants attended the first session believing they were paired with another naïve participant but in fact were paired with a confederate. Participant and confederate were briefed together by the experimenter and told they would take part in a memory study, with the second session taking place the next day. Participants were instructed that would have to act out a set of 15 shapes (=, A, C, F, H, I, J, K, L, O, P, T, V, X, △), with any part of their body or combination of body parts. The experimenter then demonstrated six different ways a shape can be created with the entire body or combination of body parts for a shape cue not used in the experiment (U). Participants were cued with a printed label of each shape. Members of the pair took turns generating actions for each cue, interleaving performing and observing actions such that performing an action in response to a cue was followed by observing the other person perform an action in response to the same cue. Each participant generated a total of 3 actions per cue, resulting in 45 performed and 45 observed actions overall. Participants were told to observe their partners during partner-generation to avoid duplicating exemplars that had already been created for a cue. Illustrations of the actions are shown in Appendix A.

Confederates (n=5) were briefed in full about the experiment prior to their participation. They were shown up to fifteen ways each shape could be made and were instructed to avoid duplicating the participants' actions.

The naïve participants returned a day later for the test phase. Participants were shown the 15 shape labels one at a time in random order. The Recall-performed group (N=18) were asked to re-perform the actions they had performed themselves and were warned not to retrieve actions they saw the other person perform. The Recall-observed group (N=17) was asked to re-perform the actions they had observed their partner perform and were warned not to perform actions they had generated themselves. Participants were asked to re-perform as many exemplars from the appropriate source (performed actions or observed actions) as they could remember for each of the shape cues, working at their own pace.

Action coding

Photographs were taken of all actions performed during generation and test for both participant and confederate to allow coding of the actions. These were coded by the experimenter (NL), using a coding scheme developed in a pilot study. For each shape between 20 and 40 distinct solutions were identified and assigned categorised numbers. The generation and retrieval phase were then coded separately. To test the reliability of the coding scheme, a subset of the photographs was coded by two independent raters naïve to the purpose of the study and experimental condition. The independent raters coded the photos for the first twenty participants, with one rater coding generation phase photographs from the first half and retrieval phase photographs from the second half of those twenty participants, while the other rater coded generation phase photographs from the second half and retrieval phase photographs from the first half of participants. Inter-rater agreement between the experimenter and the two raters was 87% and 91% each, confirming the reliability of the coding scheme. Subsequent analyses were solely based on the experimenter's judgements.

Analytic approach to address guessing

In the free recall test at retrieval, each action retrieved by participants was coded as a correct recall or a source error (the action was from the correct or incorrect source for the task respectively) or an intrusion error (the action was not generated at encoding and therefore neither seen nor performed). I will report and discuss the conventional analyses of the effect of manipulations on the frequency of correct responses, source errors and intrusion errors, with source errors the focus of our interest. However, one concern about source errors in any memory retrieval task is that source errors might either be a genuine memory error (the measure of interest) or simply a guess – an ad-hoc solution generated during the retrieval task – that just happened to be an item also generated at encoding. Source errors (i.e., false source responses to items correctly recognized as old) are therefore often analysed in relation to false source responses to novel items (i.e., false source responses to items falsely recognized as old) by measuring either the difference or the ratio of both types of errors.

Lindner et al. (2010) were interested in the specific effect additional observation had on shifting participants' response at source test to falsely respond 'performed' to items they had not performed themselves. To show that the observation inflation effect was not just an effect of guessing or response bias, they contrasted the proportion of false 'performed' responses after observation in the second encoding phase with the proportion of false 'performed' responses for actions that had not been presented in the second encoding phase. Their critical measure was therefore the difference of false 'performed' responses when those items had been additionally observed in contrast to when they had not been observed. False 'performed' responses to actions that had not been observed provides the baseline of participants giving false 'performed' responses irrespective of observing someone else perform those actions. The true effect of observation in their metric is therefore the additional proportion of false 'performed' responses observation results in beyond the basic guessing error. However, this metric cannot be easily transferred to our recall task given the total number of responses at recall differs by participants and observation is the sole encoding instance of an action, rather than additional manipulation. I therefore developed a critical measure that would similarly take accidental guessing into account and look at an effect of observation beyond that, in a conceptual replication of Lindner et al. (2010)'s metric.

I used a Monte Carlo procedure to simulate how many source errors participants would commit if they were guessing and had just generated potential shapes for each shape “on the fly” during the test phase, rather than genuinely retrieving them from what they had previously either seen or performed. The simulation was based on the distribution of actions generated by participants at encoding in response to the shape cues. I simulated the test phase of the experiment for each participant and each shape separately to take into account differences between individual participants, differing frequency profiles for the different shapes, and the typicality of individual items. To achieve this, I used as much of the participant-provided observed data as possible to ensure that the only simulated part of the experiment would be the test phase.

As a first step of the simulation process, I determined frequency norms for the different actions generated for each of the 15 shapes used in the experiment, from all participants who took part in the encoding phase. Participants generated between 20 and 40 different ways of performing each shape across the experiment, with some actions produced more frequently than others. For each shape, I converted those frequency profiles of the different actions into probability distributions, reflecting the relative probability that a particular action was produced for a given shape. For each shape, the probabilities summed to 1 to represent the entire action space.

I next used these distributions in the observed test phase for each participant. To simulate a participant’s performance I took the total number of actions they performed (i.e., reported at recall) for each shape in the test phase and randomly selected this number of actions from the overall probability distribution for that shape. This sampling was done without replacement to match the experimental procedure of only retrieving an item once. This provided us with an estimate of which actions would most likely be chosen by a participant if the participant had just generated novel solutions at test, i.e., guessed a number of unique items without memory, under the assumption that these novel solutions at test would follow the same frequency distribution as during the encoding phase. I then estimated how many of these novel (simulated) solutions matched this participant’s self-performed actions, matched the actions they observed their partner perform, or were neither seen nor performed by this participant. This provides the baseline for how self-performed, observed and novel actions should be retrieved if participants were just guessing. I repeated the sampling procedure 500 times for each participant to arrive at stable estimates. As with the observed performance, I summed the simulated performance across all shapes. Source errors were now novel (simulated) solutions that happened to have been generated by the partner in the Recall performed and self-generated actions in the Recall observed condition.

To estimate how many of the observed source errors were the product of guessing I had to scale the simulated performance to the observed performance, based on the number of intrusion errors (actions that were not generated at encoding) committed by participants in the test phase. The assumption here is that intrusion errors must be the result of guessing (e.g. based on how typical or common the actions are), because those actions do not contain source-specifying information. I created the ratio of simulated source errors over all simulated errors for the simulated data for each participant (Equation 2.1) and applied that ratio to that participant's data from the experiment (Equation 2.2) to estimate how many source errors I would predict to observe if the participant was guessing given the number of intrusion errors that particular participant committed in the test phase. This gave me the number of predicted source errors, i.e., an estimate of the number of source errors I expected if recall was based on the probability of the individual actions, in addition to the number of source errors from the experiment (that I will refer to as experimental source errors). An alternative approach would be to compare the source error ratios directly for equivalent conclusions, but the present methods is more intuitively understandable.

$$r_s = \frac{1}{500} \sum_{i=1}^{500} \frac{E_s}{E_s + N_s} \tag{2.1}$$

where:

r_s = error ratio of simulated data
E_s = frequency of source errors in simulated data
N_s = frequency of novel items in simulated data

$$E_p = \frac{N_o * r_s}{1 - r_s} \tag{2.2}$$

where:

E_p = predicted frequency of source errors
N_o = observed frequency of novel items
r_s = error ratio of simulated data

Table 2.1. Frequency of responses in Experiment 1 out of a maximum of 45

Responses	Recall performed	Recall observed
Correct	23.78 (3.81)	15.94 (3.99)
Source Errors	8.33 (3.50)	7.65 (5.30)
Intrusion Errors	6.33 (2.54)	8.29 (3.14)

Note. Mean with Standard Deviations in brackets.

At the end of the simulation process I therefore had two source error rates for all participants in the experiment: the rate predicted from guessing, and the rate that was observed in the experiment (experimental). The subsequent analyses I performed were based on these data, with Data origin (experimental, predicted) used as a factor, i.e., in Experiment 1 I can ask whether source errors (for Recall performed and Recall observed tasks) exceed the frequency I would expect if participants were just guessing. Given our theoretical questions, I will focus on that aspect of the data but will discuss the conventional analyses of the data for a complete account of the experimental results.

2.4.2 Results

Encoding Phase

While participants were instructed to avoid duplicating their own or their partner's actions during generation, some participants still committed such errors. On average, participants duplicated 0.51% (SD=1.22%) of the actions they had already performed themselves and 1.27% (SD=1.81%) of actions performed by the confederate. Confederates never duplicated their own actions and mistakenly duplicated on average 0.57% (SD=1.25%) of the participant's actions. Partner-duplicated actions were removed from the experiment, and all subsequent analyses were restricted to actions that had only been performed by one person.

Test Phase

Actions that participants retrieved in the test phase could be classified as correct responses, source errors and intrusion errors, entirely novel actions that had not been generated by either participant or confederate in the first session. The mean frequency of participants' responses is in Table 2.1. Participants' performance at retrieval was analysed with multiple 2 Task (Recall performed, Recall observed) ANOVAs separately for correct responses, source errors and intrusion errors. Overall, more items were correctly recalled in the Recall performed than Recall observed task, $F(1,33) = 35.29, MSe =$

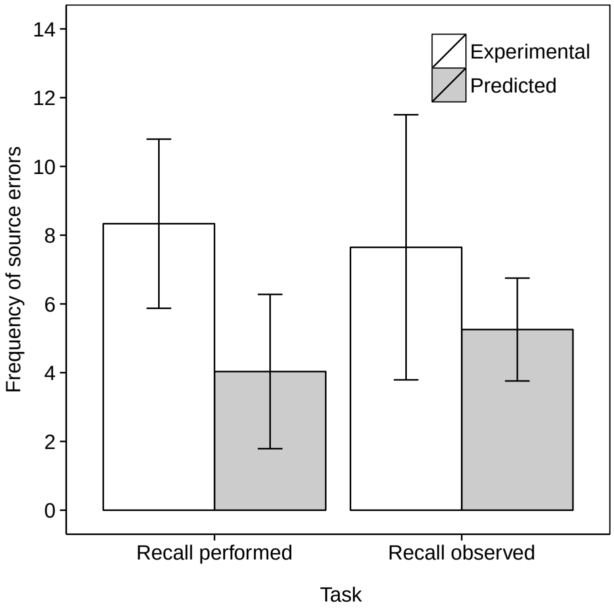

Figure 2.1. Frequency of observed and predicted source errors in Experiment 1 for both the Recall performed and Recall observed retrieval task. The error bars are 95% within-subjects confidence intervals

$15.21, p < .001, \eta_p^2 = .517$. There was no evidence for a difference in the number of source errors reported, $F < 1$, but novel intrusions occurred more often in the Recall observed than Recall performed task, $F(1,33) = 4.15, MSe = 8.11, p = .050, \eta_p^2 = .112$.

Because of my theoretical focus on source memory, I contrasted source errors committed in the experiment with those predicted if participants were guessing, based on the number of intrusion errors participants committed. Experimental and predicted frequencies of source errors are illustrated in Figure 2.1. Analysing those data as a 2 Task (Recall performed, Recall observed) x 2 Data origin (Experimental, Predicted) mixed ANOVA with repeated measures on the second factor, showed no main effect of Task, $F < 1$. However, there was a main effect of Data origin, $F(1,33) = 5.64, MSe = 12.53, p = .024, \eta_p^2 = .146$, so source errors were observed more frequently than predicted by guessing. There was no significant interaction between Task and Data origin, $F(1,33) = 1.03, MSe = 10.80, p = .32, \eta_p^2 = .030$. In subsequent step-down analyses, I tested whether observed source errors were greater than predicted in both retrieval tasks. Observed frequencies significantly surpassed predicted frequencies in the Recall performed task, $t(17) = 3.92, p < .001$, $d_{av} = 1.28$ and the Recall observed task, $t(16) = 1.84, p = .041$, $d_{av} = 0.59$, both one-tailed. This shows that people both gave away actions and claimed observed actions as self-performed.

2.4.3 Discussion

I successfully conceptually replicated the observation inflation effect of Lindner et al. (2010) in a new paradigm. Using a single encoding phase and non-verbalisable actions, I found false free-recall of observed actions as being self-performed, ruling out that the effect merely reflects an enhanced verbal or object based encoding of the seen actions or is unique to recognition tasks. Importantly, however, I also observed the opposite effect: the tendency to attribute self-performed actions to a partner. This is in line with reports of participants giving away ideas in the verbal domain (Hollins et al., 2016, 2015; Perfect et al., 2009) and action domain (Hornstein & Mulligan, 2004; Leynes & Kakadia, 2013; Rosa & Gutchess, 2011).

This latter effect argues against a simple motor activation account of observation inflation, according to which no confusion should arise for self-performed actions because both visual and motoric memories indicate self-performance. However, looking at both the numerical difference (and error bars) in Figure 2.1 and the achieved power given the effect size (d_z) shows that the result in the Recall performed task is sufficiently precise with an achieved Power = .98 (calculated with G*Power; Faul, Erdfelder, Lang, & Buchner, 2007). Interpretation of the results in the Recall observed task (achieved Power =.55) is more difficult. Given the clarity of the effect for one error but not for the other would be in line with predictions of Lindner et al. (2010) of motor activation leading to false memories of self-performance while the reverse error could be to a greater degree an artefact of guessing processes. This means that while the reverse error may clearly exist for verbal material (Hollins et al., 2016, 2015), it may be weaker for actions.

Alternatively, the lack of precision in the Recall observed task may emerge from using confederates. Confederates were encouraged to use a variety of body parts to avoid participants being able to distinguish performed and observed actions simply on the basis of the confederate, for example, only using their arms to perform shapes. Given confederates were shown actions beforehand and took part in the experiment repeatedly, their performance of the actions may have been much clearer in achieving the shape in a more prototypical way than naïve participants' performance. In fact, some naïve participants did report when they were debriefed that they had been aware they had been paired with a confederate on the basis of how the confederate performed the actions. It is possible that knowledge of the confederates or the quality of confederates' performance would influence the strategies at retrieval and thereby patterns of retrieved actions. When looking for self-performed actions, observed actions that were performed quite clearly and confidently may be more strongly represented in memory and intrude as source errors. When retrieving actions participants observed the confederate perform, their own self-performed actions do not stand out to the same degree, and participants are less likely to misattribute their own action attempts as having been performed by the confederate.

Taken altogether, Experiment 1 establishes a recall paradigm for the exploration of source confusion in action memories. Source confusion of action memories was observed with non-object-directed actions. While participants clearly misremember observed actions as self-performed, the reverse error of misremembering self-performed actions appears to occur to the same extent.

2.5 Experiment 2

While participants misremembered performed actions as observed, there is the possibility that this merely the result of using confederates. As a fundamental change in the experimental design for Experiment 2, I paired two naïve participants. If Experiment 2 shows a similar asymmetry, I would have to assume that the difference in effect size and power is a function of the retrieval task (i.e., recalling self-performed and observed actions). In contrast, if the difference emerged from the potentially biased action production of the confederates, it should now be eliminated. The control or baseline condition in Experiment thus serves as the replication of Experiment 1.

Since I established that I can elicit source confusion of action memories in this paradigm, I turned to testing a motor activation account more directly. The critical prediction of a mirroring account is that participants' motor systems 'resonate' with the actions they observe and produce an internal replica of the action, as if it were self-performed (Craighero, Bello, Fadiga, & Rizzolatti, 2002; Craighero, Fadiga, Rizzolatti, & Umiltà, 1999). In other words, encoding the motor trace during observation is the critical element responsible for source confusion. Taxing the motor system with a secondary motor task should therefore prevent the formation of such traces (cf. Bach, Allami, Tucker, & Ellis, 2014; Hamilton, Wolpert, & Frith, 2004; Vetter & Wolpert, 2000; Zwickel, Grosjean, & Prinz, 2007). Those memories now lacking the mirrored motoric information should be less likely to be confused for self-performed actions.

Thus, in Experiment 2, I asked participants to execute motor actions that differed from the actions they observed, as they observed actions. This motor performance should interfere with the encoding of motor representations of observed actions, leaving only the visual-perceptual component of the memory trace. Without the assumed mirrored motor component, participants should report fewer partner actions falsely as their own. In contrast, no such difference should be observed if the effects emerge from general source confusion processes outside the motor system.

I used two types of motor system load (in addition to the baseline/no-load control condition). First, a motor execution load task was used to directly engage execution-related motor resources. Participants were asked to walk in place, swinging their arms, as they watched their partner perform an action. These whole body movements should interfere with the generation of any motor representation of the observed action, irrespective of the body part(s) used. Similar automatic motor tasks concurrent to or immediately after encoding have been shown to disrupt subsequent recall of action phrases (Helstrup, 2001; Saltz & Donnenwerth-Nolan, 1981; Smyth et al., 1988) and the acquisition of motor skills during mental practice and imitation learning (Bach, Allami, et al., 2014).

Second, I used an action planning load task to engage higher-level action planning resources. I asked participants to remember Corsi-block sequences whilst they watched their partner perform an action. In the Corsi-block task, the experimenter taps a spatial path on a random sequence of blocks arranged on a board. The participant is then asked to reproduce the sequence of taps in the same order. To estimate a participant's span, the length of the sequence increases until the participant is no longer able to repeat the sequence in the correct order. The length of sequence participants last produced correctly is commonly referred to as participants' visual-spatial working memory span (Milner, 1971). Action planning is typically assumed to rely on such a visuospatial encoding of the action one intends to perform (Hesse & Franz, 2009; Hommel, Müsseler, Aschersleben, & Prinz, 2001). It has been argued that mirror neuron activation might not reflect only a motoric encoding of the actions, but also such planning processes. Neuroimaging studies show activation during movement planning in the prefrontal, posterior parietal cortex and premotor cortex (Hanakawa, Dimyan, & Hallett, 2008; Ikkai & Curtis, 2011), the latter two classical regions implicated in mirror neuron activation (Iacoboni et al., 2001; Koski, Iacoboni, Dubeau, Woods, & Mazziotta, 2001).

To test for the possibility that action recall and confusability emerged from the encoding of such action planning (rather than low-level motor) traces of the actions, I asked participants to remember Corsi-block sequences whilst they watched their partner perform an action. The participant was asked to reproduce the sequence after observing their partner perform an action, meaning the participant had to to encode the spatial path tapped by the experimenter as an action intention for later reproduction. Since intentions for future action production are encoded motorically (Brandimonte & Passolunghi, 1994; Freeman & Ellis, 2003; Gallivan, McLean, Valyear, Pettypiece, & Culham, 2011), I would expect the motor system to be occupied with that action plan for future performance during observation of partner's actions. Hicks, Marsh, and Cook (2005) showed that keeping two intentions in mind if both are considered manageable (here: observation of partner's performance and Corsi-block task) can lead to interference. The interference I specifically expect is one of motoric encoding.

If the assumption of mirror neuron network involvement in observation inflation is correct, either or both types of concurrent motor system activity should reduce mirroring of observed actions and subsequently reduce the number of observed actions falsely recalled as self-performed.

As a direct consequence of the theoretical predictions, the experimental design in Experiment 2 was unbalanced. Concurrent load was only directly applied to observed actions, not performed actions. Since concurrent load was applied to blocks of trials, nominally there will be performed actions encoded in Action planning load or Motor execution load blocks, but self-performance of actions always took place without a concurrent load. Any effects of concurrent load on performed actions in those blocks can therefore not be directly an effect of concurrent load but may be an effect of, for example, encoding-context or attention. Given the imbalance in the design, I will first report the full analysis, looking at all trials, performed and observed actions, in the concurrent load blocks before specifically looking at the subset of data I manipulated directly and have theoretical predictions about.

2.5.1 Method

Participants

40 members of the public participated for payment of £12. Three participants did not attend all sessions and their data were excluded from the analysis.

Procedure

Participants attended the first session in pairs. Prior to the experiment, each participant's individual Corsi-block span was assessed. The length of the tapped sequence was increased up to the point that participants failed to correctly reproduce the sequence twice. Participants' span was the maximum length of sequence they successfully reproduced twice. Participants were given the same instructions as in Experiment 1 to create exemplars for 15 shape cues with any part of their body or combination of body parts. I asked participants to create 4, not 3 exemplars as in Experiment 1, to compensate for the addition of concurrent load conditions. The 15 cues were split into 3 blocks of five cues each, with a concurrent load (Action planning, Motor execution) added to the action observation trials for two of those blocks (assignment of cues to concurrent load conditions and order of those conditions was counterbalanced across participants), and no load to the remaining block. Participants now performed 20 and observed 20 actions in each of the 3 concurrent load conditions. In the Action planning load condition, participants were shown a Corsi-block sequence at their span prior to observing their partner perform an action, then asked to reproduce the sequence after the observation. In the Motor execution load condition, participants were asked to walk in place, with exaggerated movement of both arms and legs, as they observed their partner. Performance of own actions always took place under no load.

Participants returned to retrieve either their own or their partner's actions the next day, with the retrieval task identical to Experiment 1. Responses were scored and analysed as in Experiment 1.

2.5.2 Results

Encoding phase

Concurrent load had no impact on the tendency for participants to repeat their own actions (No load: M=5.13%, SD=5.75%; Action planning load: M=5.5%, SD=7.42%; Motor execution load: M=6.35%, SD=6.11%), $F < 1$. However, concurrent load did influence the tendency to duplicate a partners actions (No load: M=9.73%, SD=6.25%; Action planning load: M=13.78%, SD=7.75%; Motor execution load: M=12.16%, SD=6.93%), $F(2,72) = 3.37, MSe = 1.83, p = .040, \eta_{\text{p}}^2 = .086$, (with Bonferroni-adjustment, none of the individual pairwise comparisons differed significantly). For the analysis of retrieval performance, only those items that had only been performed by one of the participants in the

Table 2.2. Frequency of responses in Experiment 2 out of a total of 20 per condition

	Concurrent load		
Responses	No load	Action planning load	Motor execution load
Recall performed			
Correct	6.89 (2.56)	6.37 (2.39)	6.37 (2.06)
Source Errors	2.63 (1.54)	2.26 (1.37)	2.58 (1.84)
Intrusion Errors	2.42 (1.89)	2.68 (1.83)	2.37 (1.64)
Recall observed			
Correct	4.72 (2.24)	3.50 (1.29)	3.78 (2.53)
Source Errors	3.39 (2.00)	2.83 (1.82)	3.22 (1.83)
Intrusion Errors	2.44 (1.79)	3.33 (2.83)	2.89 (2.42)

Note. Mean with Standard Deviations in brackets.

pair were included.

Test phase

I will first look at the data in its entirety as in Experiment 1. Table 2.2 shows performance across both retrieval tasks and concurrent load conditions. Participants generated fewer actions per condition in Experiment 2 than in Experiment 1, making the direct comparison of the absolute frequencies difficult. Despite the discrepancy in absolute frequencies, the rates with which self-performed and observed actions are retrieved as source errors are roughly equivalent the rate with 19% of partner actions and 17% of self-generated actions reported as source errors in Experiment 1 compared to 15% and 19% respectively in Experiment 2.

As in Experiment 1, performance was analysed as multiple two-way ANOVAs with factors of Task (Recall performed, Recall observed) and 3 Concurrent load (No load, Action planning load, Motor execution load) with repeated measures on the second factor separately for correct responses, source errors and intrusion errors in the first instance.

Participants reproduced fewer correct responses in the Recall observed than the Recall performed task, $F(1,35) = 17.64, MSe = 3.39, p < .001, \eta_{\mathrm{p}}^2 = .335$, replicating Experiment 1. There was a main effect of Concurrent load, $F(2,70) = 3.49, MSe = 2.34, p = .036, \eta_{\mathrm{p}}^2 = .091$, with bonferroni-adjusted pairwise comparisons not showing a significant difference between individual concurrent load conditions. Collapsing across the two load conditions showed that correct recall was lower when actions were observed under load compared to no load, $F(1,35) = 6.29, MSe = 23.95, p = .017, \eta_{\mathrm{p}}^2 = .152$. There was no significant interaction between retrieval task and concurrent load for correct responses, $F < 1$.

For source errors, there was no significant effect of retrieval task, $F(1,35) = 2.10, MSe = 1.90, p = .16, \eta^2_\text{p} = .057$, or concurrent load, $F(2,70) = 1.27, MSe = 1.7, p = .29, \eta^2_\text{p} = .035$, nor was there a significant interaction, $F < 1$. Similarly, intrusions errors did not show main effects of Task, $F < 1$, or Concurrent load, $F(2,70) = 1.41, MSe = 2.22, p = .25, \eta^2_\text{p} = .039$. There was no significant interaction, $F < 1$.

Since concurrent load was only applied during the observation of partner's actions, I will report the effects of concurrent load on those subsets of data – correct responses in the Recall observed task and source errors in the Recall performed task – separately here. For correct responses in the Recall observed task, concurrent load had a marginally significant effect, $F(2,34) = 3.17, MSe = 2.33, p = .055, \eta^2_\text{p} = .157$, with Bonferroni-adjusted comparisons not showing a significant difference between conditions. For source errors in the Recall performed task, concurrent load did not lead to a significant effect on source errors in the Recall performed task, $F < 1$.

In sum, what I have shown so far is that concurrent load decreases correct recall of observed actions, but has no significant effect on false recall of observed actions in the Recall performed task.

As in Experiment 1, the aspect of the data I am primarily interested in are source errors. In order to test the claim of the motor simulation account that more source errors would be made when recalling self-performed rather than observed actions, I looked at the number of source errors in the experiment relative to the number prediced by guessing. Since concurrent load was only manipulated when participants observed their partner's actions, but not when they performed them, the Action planning load and Motor execution load conditions were not equivalent across retrieval tasks. I therefore only used the No load condition in this comparison (see Figure 2.2).

A 2 Task (Recall performed, Recall observed) x 2 Data origin (Experimental, Predicted) mixed ANOVA with repeated measures on the second factor did not reveal a significant main effect of Task, $F < 1$. As in Experiment 1, experimental source errors exceeded those predicted by guessing, $F(1,35) = 30.46, MSe = 1.67, p < .001, \eta^2_\text{p} = .465$. The interaction was not significant, $F(1,35) = 1.61, MSe = 1.67, p = .21, \eta^2_\text{p} = .044$. Stepdown analyses showed that observed errors surpassed predicted errors in both retrieval tasks, but the effect was smaller in the Recall own task, $t(18) = 3.15, p = .003, d_{av} = 0.98$, than the Recall partner task, $t(18) = 4.58, p < .001, d_{av} = 1.26$, both one-tailed. This means that

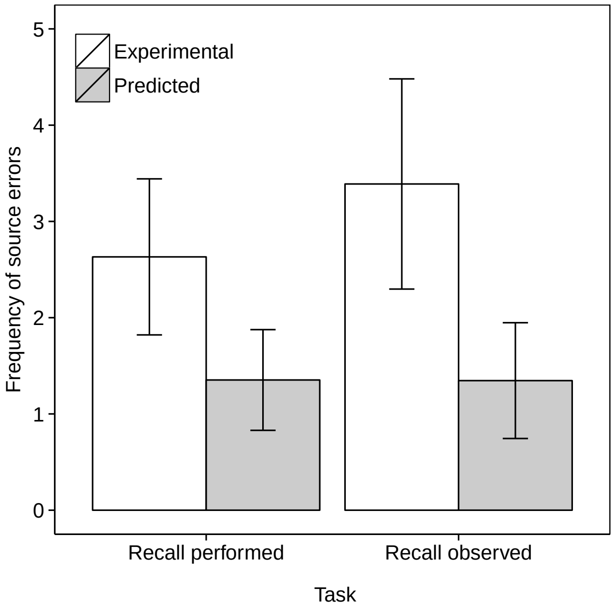

Figure 2.2. Frequency of observed and predicted source errors in Experiment 2 in the No load condition in the Recall performed and Recall observed task. The error bars are 95% within-subjects confidence intervals.

the higher rate of errors in the Recall performed task in Experiment 1 was not replicated here with naïve participants and showed, if anything, the opposite pattern.

I finally tested whether concurrent (Action planning or Motor execution) load would reduce source errors in the Recall performed task. The results are shown in Figure 2.3. I analysed the data for the Recall performed group with a 2 Data origin (Experimental, Predicted) x 3 Concurrent load (No load, Action planning load, Motor execution load) repeated measures ANOVA. As before, source errors were observed more frequently than predicted, $F(1,18) = 22.57, MSe = 1.77, p < .001, \eta^2_{\mathrm{p}} = .556$. There was no evidence for an effect of Concurrent load nor was there a significant interaction, both Fs< 1. In subsequent step-down analyses, I tested whether observed source errors were greater than predicted in every concurrent load condition. Observed frequencies significantly surpassed predicted frequencies in the Control, $t(18) = 3.15, p = .003, d_{av} = 0.98$, Action planning load, $t(18) = 2.73, p = .007, d_{av} = 0.81$, and Motor execution load condition, $t(18) = 2.55, p = .010, d_{av} = 0.91$, all one-tailed.

I obtained null effects for Concurrent load in our comparisons. A null-result in frequentist Null-hypothesis significance testing (NHST) does not allow any theoretical conclusion about the status of the tested theories since a high p-value may reflect either that the data is insensitive or may reflect a genuine rejection of a predicted (or any) alternate theory. In a frequentist approach, in short, there is no way of telling what a particular

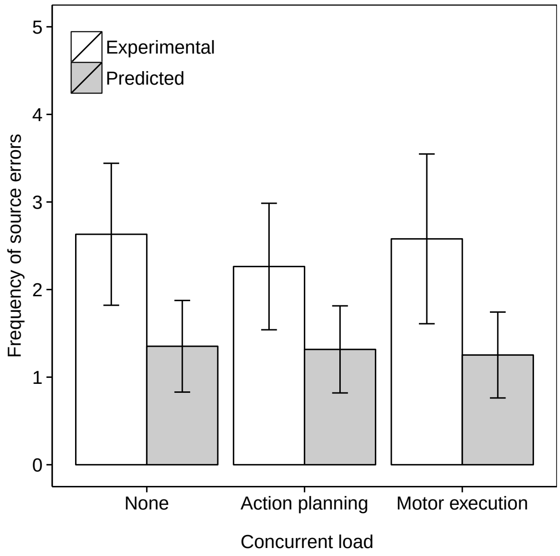

Figure 2.3. Frequency of observed and predicted source errors in Experiment 2 in the Recall performed task across all concurrent load conditions. The error bars are 95% within-subjects confidence intervals.

non-significant result means. Bayesian analysis, in contrast, allows comparative estimation of evidence in favour of two theories, thus allowing conclusions to be drawn about the status of the theories that were tested (Dienes, 2008). The relative strength of evidence for two theories (for example a null- versus an alternate theory) tested against one another is indicated by the Bayes factor that can vary between 0 and ∞, where values below 1 indicate evidence for the null-hypothesis, and above 1 in favour of the alternate hypothesis. The closer to 1, the less sensitive the evidence is. In a cut-off suggested by Jeffreys (1961), a Bayes factor greater than 3 or less than 1/3 indicates substantial evidence for the alternate and null-hypothesis respectively, while anything between those numbers only indicates anecdotal evidence.

I tested the effect of Concurrent load using the *BayesFactor* package in R (Morey & Rouder, 2015) with the *AnovaBF* function. I defined the variables as in the ANOVAs reported above.

I tested the effect of 2 Data origin (Experimental, Predicted) x 3 Concurrent load (No load, Action planning load, Motor execution load) in the Recall performed task. The winning model was the one with a main effect of Data origin with a Bayes factor of 23121. This means it the data provides $23 * 10^3$ times the evidence in favour of an effect of Data type than the null. Concurrent load specifically results in a Bayes factor of 0.09, this means that the data provides 10x more evidence for the null effect than for an effect of Concurrent load.

Given this analysis takes into account both experimental and predicted source errors (with predicted source errors largely invariant to the manipulation), I finally specifically looked at the effect of Concurrent load on experimental source errors in the Recall performed task. Even when excluding predicted source errors, the data provides 5.5 times more evidence for no effect of Concurrent load than for an effect of Concurrent load.

Rather than simply saying that there is no evidence for an effect of Concurrent load on source errors, in fact, there is evidence for the absence of an effect of Concurrent load on false retrieval of observed actions as self-performed.

2.5.3 Discussion

Experiment 2 showed clear source errors in both the Recall performed and Recall observed task. If anything, source errors were more frequent in the Recall observed than the Recall performed task. This suggests that the weaker effect in the Recall observed task in Experiment 1 may have been due to the use of confederates and did not directly reflect anything about the mechanism leading to source confusion of actions.

This pattern of bidirectional source errors in Experiment 2 is not consistent with a mirror neuron account of observation inflation. Under this account, observation of actions leads to motoric encoding of those actions. At retrieval, this motor trace is taken as an indication of the action having been self-performed since motor traces always indicate performance of actions. In this account, the reverse error of mistaking performed actions as observed cannot be explained. Thus, motor traces generated during action observation would only predict errors in Recall performed, not the Recall observed task.

In Experiment 2, I directly tested the role of the motor trace during observation of actions for subsequent source confusions. I predicted that motor load concurrent to watching someone perform actions should lead to fewer of those actions being falsely retrieved as self-performed, as own motor execution should interfere with generating motoric memory traces. In fact, neither motor execution or action planning concurrent to observation decreased source errors, and source errors surpassed predicted frequencies in each load condition.

Since I did not observe an effect of concurrent load on source errors, it is possible that our manipulation simply did not affect the encoding of observed actions at all. However, I did see that concurrent load increased partner-duplications at generation and decreased correct recall of partner actions. This means, while there was no effect of the manipulation on source errors in the Recall performed task, the manipulation did affect participants' performance in the experiment overall.

A second possibility is that the free report task I chose as the memory test cannot clearly reflect an effect of the manipulation on source confusion specifically. To make source judgements in a free recall task, participants first have to generate an action and then decide to report or withhold that action from report. Here participants' report conflates generation of solutions at test and source decisions about the generated solutions. Both of these aspects of the retrieval task can be separately affected by experimental manipulations. In a recognition test, on the other hand, the items at test are presented by the experimenter and participants' performance clearly reflects their source decision. Here, participants do not need to generated the solutions before making their decision. The results in a standard free recall task are now the result of both of these processes. It is possible that participants in a free report simply neglect the source of retrieved actions and report all actions that come to mind without engaging in explicit monitoring of the source.

Since I am interested in the effect of the motor trace for the source confusion of performed and observed actions, specifically, I will adapt our free report task to separate out the source decision from the item generation dimension in the free report task.

2.6 Experiment 3

Experiment 2 showed that concurrent motor load affected correct retrieval of observed actions but not false retrieval of observed actions as self-performed. It is possible that asking participants to freely report self-performed actions conflated retrieval of those actions and decisions about the source-appropriateness of actions generated at report (though admittedly overall decrease of memory strength and increase in source strength should both lead to a reduction in source errors). Since the predictions of a mirror neuron network account are about the source judgements, I decided to move to a task that allows looking at both aspects of recall separately.

The retrieval task in Experiment 3 was therefore separated into two separate stages, following an extended recall procedure developed by Bousfield and Rosner (1970), and more recently used by Kahana et al. (2005) and Hollins et al. (2016). Here participants are asked to make an explicit source judgement at test, so any source errors that arise cannot be explained by the kind of source neglect mentioned above. I asked participants to perform all actions that came to mind at test given a shape cue and retrieval task. For each action performed, participants were explicitly asked to consider its source carefully and to decide whether or not it was compliant with their retrieval goal (i.e. to recall performed or observed actions). The actions that participants reported are therefore only those they had explicitly attributed to the required source.

Rather than looking only at the actions participants choose to report as self-performed (or observed), I can now look separately at all actions generated by participants at retrieval and the source decisions they made for those actions. I will refer to those separate aspects (implemented here as separate phases of the retrieval process) as availability for the generation of items at retrieval and monitoring for the source decisions made about those items (Jacoby, Kelley, and McElree (1999) used the terms early-selection and late-correction respectively). Participants' performance in the availability phase of the retrieval task will reflect which actions are available to participants at retrieval given the retrieval task (that is, what is the size of the search set they open when cued by a shape and the task to recall performed or observed actions). Monitoring refers to participants deciding whether the candidate actions they generate do in fact comply with their retrieval task. An effect of concurrent motor load on source confusion should enable participants to be more accurate in this monitoring process and be more likely to withhold observed actions generated as candidate actions from report as self-performed.

Additionally, I replaced the Motor execution load from Experiment 2 with a Verbal load task to be able to pinpoint a specific motor load effect compared to a generic cognitive one. If source errors are only due to mirroring processes, then I would expect to see a reduction of false reports of partner actions as self-performed only under concurrent action planning but not concurrent verbal load. In contrast, if they emerge from a more general source, both loads should affect source errors equally if the tasks are equally demanding. To ensure that the tasks are equally demanding, I estimated participants' span for both concurrent load task prior to encoding.

2.6.1 Method

Participants

42 members of the public participated for payment of £12. Four participants did not attend all sessions and their data were excluded from the analysis.

Experimental procedure

Prior to the experiment, each participant's individual Corsi block-tapping span and forward digit span were assessed. The Action planning load condition was identical to Experiment 2. The Motor execution load condition from Experiment 2 was replaced by a Verbal load condition. Participants heard a sequence of digits at their individual span prior to observing their partner perform an action and were asked to reproduce the sequence after their partner completed their action. Concurrent load was only administered during observations of partner actions, not during execution of own actions. As in Experiment 2, participants generated 4 exemplars each in response to 15 shape cues split over 3 concurrent load conditions. Participants returned the next day individually for an extended recall task. They were instructed to retrieve and re-perform either their own actions or those they had observed their partner perform the previous day. They were told that, as they tried to remember their own (or their partner's) actions, other actions may come to mind such as their partner's actions when they had to remember their own actions, or entirely new ways of performing each shape. They were encouraged to perform everything that came to mind as they tried to remember their own (or their partner's) actions, and to indicate verbally for each action whether it was a target action or not. They were not instructed to search their memory for actions from both sources, nor to generate entirely new actions.

2.6.2 Results

Encoding phase

Unlike Experiment 2, there was a main effect of Concurrent load on participants repeating their own actions (No load: M=2.63%, SD=4.15%; Action planning load: M=4.87%, SD=6.09%; Verbal load: M=5.66%, SD=7.28%), $F(2,74) = 5.02, MSe = .75, p = .009, \eta_p^2 = .120$, with Bonferroni-adjusted comparisons showing that participants repeated more of their own actions in concurrent load conditions (Action planning load, $p = .033$; Verbal load, $p = .023$) compared to the No load condition. As in Experiment 2, participants more often copied partner's actions under load, but this effect was not significant here (No load: M=8.68%, SD=6.75%; Action planning load: M=10.66%, SD=7.90%; Verbal load: M=11.05%, SD=7.98%), $F(2,74) = 1.12, MSe = 2.45, p = .33, \eta_p^2 = .029$. For the analysis of retrieval performance, only those items that had only been performed by one of the participants in the pair were included. Note, verbal and motor load do not differ in their

Table 2.3. Frequency of responses in Experiment 3 out of 20 actions per condition

	Concurrent load		
Responses	No load	Action planning load	Verbal load
Recall performed			
Correct responses	8.21 (2.04)	7.21 (1.39)	7.00 (2.24)
Source errors	0.95 (0.85)	1.74 (1.59)	1.37 (1.12)
Intrusion errors	1.89 (1.52)	1.89 (1.63)	1.58 (1.77)
Recall observed			
Correct	5.21 (2.24)	3.68 (2.19)	3.37 (1.61)
Source Errors	1.37 (0.83)	2.00 (1.60)	1.68 (1.34)
Intrusion Errors	2.68 (1.45)	3.05 (2.44)	2.74 (1.56)

Note. Mean with Standard Deviations in brackets.

impact on encoding performance.

Test phase

The extended recall task allows me to look at three aspects of the free recall performance separately (final report, availability and monitoring). I will first look at the final product of participants' report, analogous to the free recall performance in Experiment 1 and 2. These are the actions participants' report as task-compliant, i.e. excludes the actions they produce at retrieval but choose to withhold from report. As in Experiment 1 and 2, I will first present the conventional analysis followed by the comparison of predicted and experimental source errors.

Final report I will again first report the conventional analyses for the actions participants report at retrieval. The means are reported in Table 2.3. Performance (Correct responses, Source errors, Intrusion errors) was analysed as a multiple 2 Task (Recall performed, Recall observed) x 3 Concurrent load (No load, Action planning load, Verbal load) ANOVA with repeated measures on the second factor. Participants in the Recall performed task recalled more items correctly overall compared to those in the Recall observed task, $F(1,36) = 41.99, MSe = 2.59, p < .001, \eta_p^2 = .538$. Concurrent load at encoding influenced the number of items retrieved correctly, overall, $F(1.707, 61.453) = 9.02, MSe = 3.28, p = .001, \eta_p^2 = .200$, corrected for violations of sphericity. Pairwise bonferroni-adjusted comparisons showed that while each of the two load conditions led to fewer correct actions being reported (both $p = .001$) compared to the No load condition, the two load conditions did not differ significantly from one another, $p = 1$. There was no significant interaction, $F < 1$.

There was no main effect of Task on source errors, $F(1,36) = 1.56, MSe = 0.68, p = .22, \eta_p^2 = .041$. Concurrent load at encoding affected the frequency of source errors, $F(2,72) = 3.52, MSe = 1.37, p = .035, \eta_p^2 = .089$, with bonferroni-adjusted comparisons showing that source errors were reported more frequently when actions were encoded under Action planning load compared to No load ($p = .047$) with Verbal load not significantly increasing source errors ($p = .43$), nor the two load conditions showing a significant difference from one another ($p = .67$). There was no interaction, $F < 1$.

Participants in the Recall observed task committed more intrusion errors than those in the Recall performed task, $F(1,36) = 8.39, MSe = 2.59, p = .006, \eta_p^2 = .189$. Load did not affect intrusion errors, nor was there an interaction, Fs< 1.

As in Experiment 2, I analysed the effect of concurrent load on observed actions separately for both retrieval tasks since only encoding of observed, and not performed actions, was directly manipulated with a concurrent load. Concurrent load had an effect on the correct retrieval of observed responses in the Recall observed task, $F(2,36) = 10.19, MSe = 1.81, p < .001, \eta_p^2 = .361$. Bonferroni-adjusted comparisons showed significantly more observed actions correctly recalled when they had been encoded under No load compared to either the Action planning load ($p < .001$) or the Verbal load ($p = .004$) condition, with no significant difference between Action planning and Verbal load conditions ($p = 1$). Concurrent load had a marginally significant effect on the frequency of source errors in the Recall performed task, $F(2,36) = 2.66, MSe = 1.11, p = .083, \eta_p^2 = .129$, with bonferroni-adjusted comparisons not significant.

As in Experiment 1 and 2, I contrasted the experimental source error rates with the frequency of source errors participants were predicted to make if they were guessing. As in Experiment 2, I first compared experimental and predicted source errors in the Recall performed and Recall observed task, when all actions were encoded without a concurrent load (see Figure 2.4). I analysed the data with a 2 Task (Recall performed, Recall observed) x 2 Data origin (Experimental, Predicted) mixed ANOVA with repeated measures on the second factor. Source errors were more likely in the Recall observed than Recall performed task, $F(1,36) = 5.27, MSe = 0.54, p = .028, \eta_p^2 = .128$. However, there was no evidence that source errors were observed more frequently in the experiment than predicted from guessing, $F(1,36) = 1.35, MSe = 0.49, p = .25, \eta_p^2 = .036$, nor was there a significant interaction, $F < 1$. Separate analyses showed that in fact observed frequencies did not significantly surpass predicted frequencies in either Recall performed, $t(18) = 0.75, p = .23, d_{av} = 0.21$ or Recall observed task, $t(18) = 0.89, p = .19, d_{av} = 0.31$, both one-tailed.

I next tested the effect of concurrent load on source errors committed in the Recall performed task only, as shown in Figure 2.5. I analysed the data as a 2 Data origin (Experimental, Predicted) x 3 Concurrent load (No load, Action planning load, Verbal load) repeated measures ANOVA. Source errors were committed more frequently than predicted, $F(1,18) = 10.16, MSe = 1.04, p = .005, \eta^2_{\text{p}} = .361$. There was no evidence for an overall effect of Concurrent load, $F(2,36) = 2.16, MSe = 0.62, p = .13, \eta^2_{\text{p}} = .107$. The interaction was not significant, $F(2,36) = 2.05, MSe = 0.83, p = .14, \eta^2_{\text{p}} = .102$. In subsequent step-down analyses, I tested whether observed source errors were greater than predicted source errors in every concurrent load condition. Observed frequencies significantly surpassed predicted frequencies in the Action planning load, $t(18) = 2.61, p = .009, d_{av} = 0.82$, and Verbal load condition, $t(18) = 2.18, p = .021, d_{av} = 0.74$, but not in the Control condition, $t(18) = 0.75, p = .23, d_{av} = 0.21$, all comparisons one-tailed.

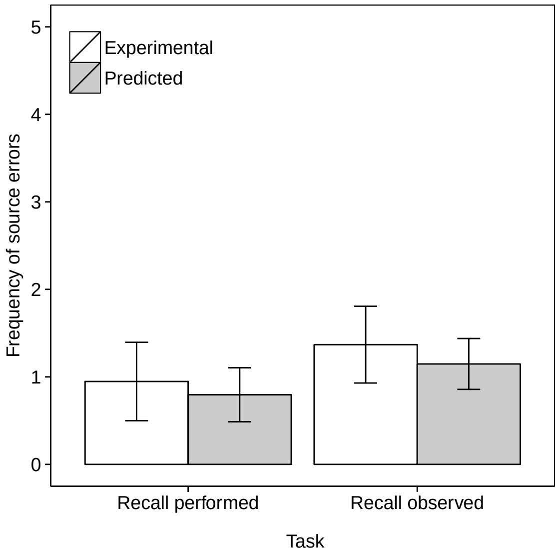

Figure 2.4. Frequency of observed and predicted source errors in Experiment 3 in the No load condition in the Recall performed and Recall observed task. The error bars are 95% within-subjects confidence intervals.

Extended Recall task The standard free report task only allows me to look at the final product of participants' retrieval as above. The extended recall task that I used in this experiment also allows me to look at two aspects of the retrieval process separately, that result in that final report: the overall retrieval of actions (availability) and what proportion of those actions participants choose to report as complying with their retrieval task

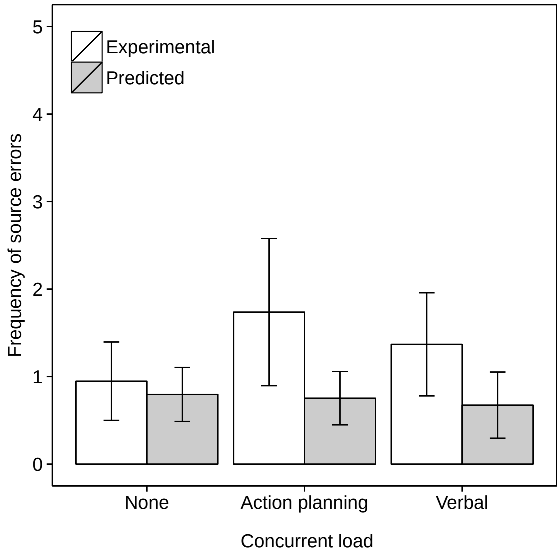

Figure 2.5. Frequency of observed and predicted source errors in Experiment 3 in the Recall performed task across all concurrent load conditions. The error bars are 95% within-subjects confidence intervals.

(monitoring). I will first look at availability, i.e. the actions participants produce at retrieval irrespective of whether they report them as task-compliant or decide to withhold them from report.

Availability The mean frequency of actions participants generated at retrieval, irrespective of their choice to report or withhold them, are in Table 2.4. The availability of actions from different sources (i.e., excluding novel actions) was explored with a 2 Task (Recall performed, Recall observed) x 2 Source (Performed actions, Observed actions) x 3 Concurrent load (No load, Action planning load, Verbal load) mixed ANOVA on frequency of responses with repeated measures on the second and third factor.

The main effect of Task was not significant, $F < 1$, but there was a main effect of Source, $F(1,36) = 86.50, MSe = 6.25, p < .001, \eta_{\text{p}}^2 = .706$, with performed actions being retrieved more frequently than observed actions. There was also a main effect of concurrent load, $F(2,72) = 5.88, MSe = 1.32, p = .004, \eta_{\text{p}}^2 = .140$. Actions generated in the Action planning load (marginally significant at $p = .054$, $d_{av} = .41$) or Verbal load ($p = .006$, $d_{av} = .59$) conditions were retrieved less frequently than those generated during the No load condition. Action planning and Verbal load condition did not differ significantly from one another ($p = 1$, $d_{av} = .15$). This suggests that the concurrent load conditions overall decreased encoding of actions in that block of trials.

The main effects were qualified by a significant Task by Source interaction, $F(1,36) = 12.42, MSe = 6.25, p = .001, \eta_{\text{p}}^2 = .256$. Unsurprisingly, in the Recall performed task, per-

Table 2.4. Extended Recall in Experiment 3: Availability

Responses	Concurrent load No load	 Action planning load	 Verbal load
Recall performed			
Performed actions	9.84 (2.40)	8.89 (2.39)	8.68 (2.55)
Observed actions	5.21 (2.20)	5.05 (2.86)	4.42 (1.51)
Novel actions	4.89 (2.74)	5.21 (2.20)	4.32 (2.83)
Recall observed			
Performed actions	7.58 (3.11)	7.74 (2.95)	7.53 (3.30)
Observed actions	7.26 (2.89)	4.95 (2.49)	4.89 (1.90)
Novel actions	5.26 (2.00)	5.58 (3.01)	5.00 (2.24)

Note. Frequency of actions generated by participants at retrieval irrespective of source decision. SDs in brackets.

formed actions were retrieved more frequently than observed ones $F(1,18) = 92.21, MSe = 5.57, p < .001, \eta_p^2 = .837$. Interestingly, in the Recall observed task, the effect was in the same direction but smaller, $F(1,18) = 15.05, MSe = 6.92, p = .001, \eta_p^2 = .455$. This pattern is in line with findings in the extended source-cued free recall task shown by Hollins et al. (2016).

All of the above effects were qualified by a significant three-way interaction, $F(2,72) = 4.01, MSe = 3.59, p = .022, \eta_p^2 = .100$. I looked at the Source by Concurrent load interaction separately for the Recall performed and Recall observed task. There was no evidence for a significant interaction in the Recall performed task, $F < 1$, but the interaction was significant in the Recall observed task, $F(2,36) = 5.79, MSe = 3.15, p = .007, \eta_p^2 = .243$. There was no significant effect of Concurrent load on generation of performed actions at retrieval, $F < 1$, but there was on observed actions, $F(1.53, 27.45) = 12.05, MSe = 2.88, p < .001, \eta_p^2 = .401$. Boferroni-adjusted comparisons showed that observed actions encoded under action planning load ($p < .001$) and verbal load ($p = .007$) were less likely to be generated while retrieving observed actions, than actions generated under no load. There was no significant difference between the action planning load and verbal load conditions ($p = 1$).

In other words, the overall main effect of Concurrent load on the availability of actions at retrieval is primarily driven by any kind of cognitive load decreasing the number of observed actions that participants are able to generate when they are looking to retrieve them.

There was no evidence that novel actions were generated more frequently at retrieval in

Table 2.5. Extended Recall in Experiment 3: Monitoring

Source	Concurrent load		
	No load	Action planning load	Verbal load
Recall performed			
Performed actions	0.84 (0.14)	0.81 (0.13)	0.81 (0.15)
Observed actions	0.19 (0.18)	0.36 (0.27)	0.30 (0.24)
Novel actions	0.39 (0.27)	0.39 (0.32)	0.32 (0.34)
Recall observed			
Performed actions	0.19 (0.12)	0.26 (0.19)	0.23 (0.19)
Observed actions	0.72 (0.17)	0.71 (0.27)	0.69 (0.22)
Novel actions	0.52 (0.26)	0.50 (0.31)	0.52 (0.23)

Note. Proportion of actions generated at retrieval from respective sources that is accepted by participants as complying with their retrieval task. SDs in brackets.

either retrieval task, $F(1,36) = 1.02, MSe = 2.09, p = .32, \eta_{\text{p}}^2 = .028$, and Concurrent load did not affect generation of novel actions at retrieval, $F < 1$. There was no significant interaction, $F < 1$.

Monitoring I next looked at the proportion of retrieved actions that participants chose to report as complying with their retrieval task. The mean proportions are shown in Table 2.5. I analysed the data as above with a 2 Task (Recall performed, Recall observed) x 2 Source (Performed actions, Observed actions) x 3 Concurrent load (No load, Action planning load, Verbal load) ANOVA, with repeated measures on the second and third factor.

There was a significant main effect of Task, $F(1,35) = 8.98, MSe = 0.04, p = .005, \eta_{\text{p}}^2 = .204$. Participants in the Recall performed task reported a higher proportion of performed and observed actions than participants in the Recall observed task. There were no significant main effects of Source, $F(1,35) = 1.14, MSe = 0.04, p = .29, \eta_{\text{p}}^2 = .032$, or Concurrent load, $F(2,70) = 1.38, MSe = 0.03, p = .26, \eta_{\text{p}}^2 = .038$. There was a significant Task by Source interaction, $F(1,35) = 374.11, MSe = 0.04, p < .001, \eta_{\text{p}}^2 = .914$, showing that participants were able to comply with their retrieval task. Participants in the Recall performed task reported a higher proportion of performed than observed actions, $F(1,17) = 162.35, MSe = 0.05, p < .001, \eta_{\text{p}}^2 = .905$, while participants in the Recall observed task reported a higher proportion of observed than performed actions, $F(1,18) = 224.41, MSe = 0.03, p < .001, \eta_{\text{p}}^2 = .926$. Neither the Task by Concurrent load nor the Source by Concurrent load interactions were significant, $F < 1$. The above effects were qualified by a significant three-way interaction, $F(2,70) = 3.26, MSe = 0.03, p = .044, \eta_{\text{p}}^2 = .085$.

I looked at the Source by Concurrent load interaction separately for the Recall performed and Recall observed task. The interaction was significant in the Recall performed task, $F(2,34) = 3.98, MSe = 0.02, p = .028, \eta_{\text{p}}^2 = .190$. There was no significant effect of Concurrent load on performed actions, $F < 1$, but there was on observed actions, $F(2,34) = 3.49, MSe = 0.04, p = .042, \eta_{\text{p}}^2 = .170$. Bonferroni-adjusted comparisons showed that the proportion of observed actions that participants included in their report was higher in the Action planning load than No load condition, $p = .018$, while the remaining comparisons did not differ significantly from one another, ps$> .36$. The Source by Concurrent load condition interaction was not significant in the Recall observed task, $F < 1$. This suggests that overall concurrent load increased the proportion of observed actions that participants falsely reported at retrieval.

Novel actions were reported to a greater degree in the Recall observed task than Recall performed task, $F(1,35) = 5.71, MSe = 0.11, p = .022, \eta_{\text{p}}^2 = .140$. There was no significant effect of Concurrent load, $F < 1$, or a significant interaction, $F < 1$.

In sum, concurrent load affected the retrieval of observed actions during both availability and monitoring phases but did so differentially in both retrieval task. When attempting to search for observed actions, observed actions were less available when they had been encoded under concurrent load compared to the baseline condition. There was no additional detrimental effect of concurrent load on the monitoring of those actions. When attempting to retrieve performed actions, on the other hand, concurrent load did not affect which observed actions came to mind but it lead participants to falsely report a higher rate of those observed actions as performed. There was no evidence that the verbal and motor-system based load differed in their effects.

2.6.3 Discussion

The final product of participants' retrieval in Experiment 3 shows a similar pattern of data to Experiment 1 and 2. Participants both falsely recall observed actions as performed and performed actions as observed. There is no evidence for errors occurring more frequently in one direction than the other. However, in contrast to the previous experiments, source errors of either kind did not occur more frequently than guessing in the control (No load) condition. As a comparison with Experiment 2 shows, the additional source-focus in the extended recall task in Experiment 3 reduced source errors in the conditions that had allowed participants to sufficiently encode actions (in the No load condition).

In addition, motor activation accounts of observation inflation predict that performing a motor task concurrent to observation of actions should decrease source errors compared to observing actions without a concurrent task. The same should not be true if the concurrent task is non-motoric. Experiment 3 did not confirm this prediction. Source errors were higher than predicted from guessing for both the motor load and the verbal load conditions and were, if anything, more frequent in the verbal load condition. While I theoretically predicted an effect of concurrent load for observed actions specifically, closer examination of the data shows similar, weaker effects of concurrent load exist for performed actions. Recall, performed actions were never encoded under load, so any effects here can only be effects of, for example, task-switching or inattention. This suggests that the effects of the concurrent load on observed actions may be in part a result of task-switching costs.

The patterns of data are somewhat more distinct when looking at both phases of the retrieval task separately. What effect did our concurrent load manipulation have on observed actions? Interestingly, the effects differ by retrieval task, with three-way interactions in both retrieval phases pointing towards the differential pattern of the concurrent load manipulation. Concurrent load impaired availability of observed actions in the Recall observed but not the Recall performed task. This means the detrimental effects of concurrent load only become apparent when participants' search set in a task requires them to search for observed actions. While they can recover observed actions when they encoded them without a load, they struggle to do so when they observed actions under a cognitive load. Observed actions are only incidentally found during the search for performed actions in the Recall performed task, so the effects of the encoding condition do not become apparent here.

In the monitoring phase the effect is reversed, with concurrent load affecting participants' willingness to report actions only in the Recall performed task. Here participants were more likely to falsely report observed actions as self-performed, a prediction that runs against predictions of the motor simulation account. It is possible that this reflects an overall change in response bias. Maybe concurrent load at encoding made participants overall more liberal in their reporting? Since correct reports of performed actions and false reports of novel actions are not affected by the concurrent load manipulation, this seems unlikely. It is more plausible that participants' encoding of the actions encoded under load was insufficient to result in correct source attribution. Observed actions in the Recall performed task are those that simply happened to come to mind as participants were looking to retrieve performed actions. Actions that had not been sufficiently encoded to allow recovery of features that would lead to correct monitoring were now more likely to be falsely reported. Interestingly, while concurrent load affected participants' retrieval of observed actions, it did not additionally affect their monitoring. It is likely that once observed actions were retrieved in the Recall observed task, participants were certain they had observed those actions, so there were no additive effects of load on accepting those actions as task-compliant.

Importantly, there was no clear difference in effect for a motor-system load or a verbal load here either. The effect of concurrent load therefore was not a selective impairment of the motor trace at encoding but potentially an overall impairment to attention that globally impaired encoding of features of observed actions. While I observed the opposite trend in the data than I would have expected under motor activation account of observed actions, it is possible that even if I impaired motor encoding and enhanced source performance, the impairment to the remainder of encoding may have counteracted this effect.

2.7 Meta-analysis of the effects in Experiment 1 through 3

The observation inflation effect (Lindner et al., 2010) is the false retrieval of observed actions as being self-performed and has been attributed to mirror neuron network activation at the time of observation. In this first series of experiments I tested 1) whether I can conceptually replicate the observation inflation effect with a simpler paradigm that rules out influences of verbal and object-based encoding of the actions, 2) whether there is a complementary, reverse error during the retrieval of partner's actions, and 3) whether motor system loads reduce the observation inflation effect. So far I have presented results separately for all three experiments. As I pointed out earlier, the imprecision is high (the power low) in some of the single data points, making conclusive interpretation of the overall effect difficult. To address my main theoretical questions, I therefore integrated the data across experiments into two forest plots (Cumming, 2013) with the expectation that this meta-analytic approach would be less affected by noise than the individual studies. Rather than looking at individual comparisons, I can now look at summary effects that contain the effects of the individual comparisons, with more precise effect size estimates contributing more to the summary effect.

With respect to my first two questions the results are clear. Figure 2.6 shows the memory error effects for the control (No load) conditions for all 3 experiments relative to guessing

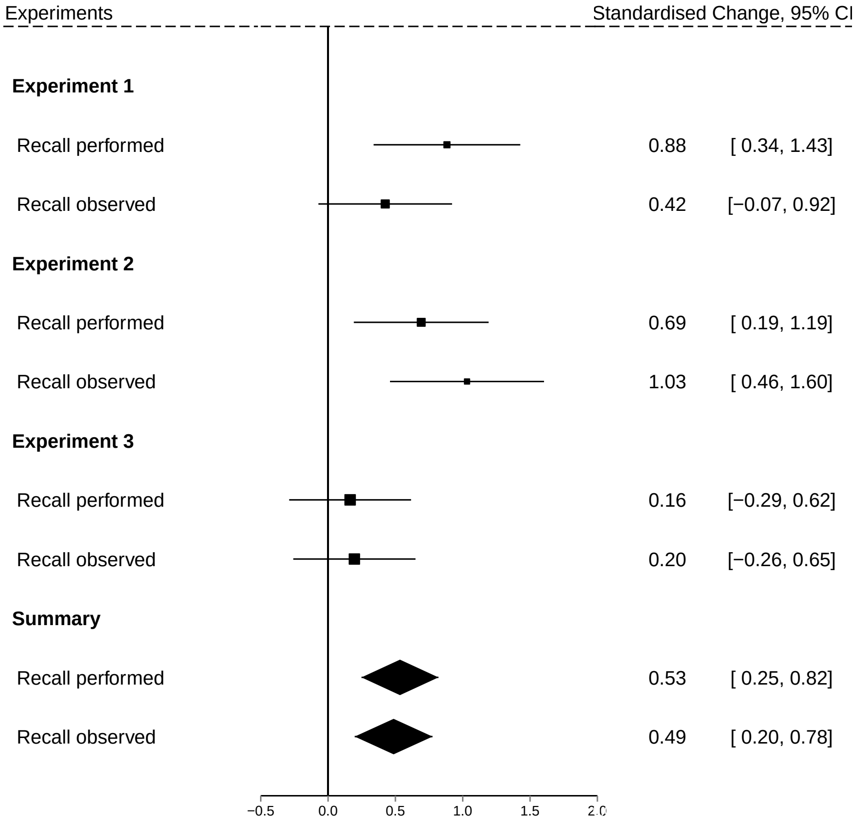

Figure 2.6. Standardized mean change between experimental and predicted source errors committed in the No load conditions in the Recall performed and Recall observed task in Experiments 1 through 3. 0 indicates source error rates as would be predicted by guessing. The error bars are 95% confidence interval. The polygons represent the summary effects (fixed effects) across the indicated conditions.

performance at 0. Summary effect sizes for the Recall performed and Recall observed task, given at the bottom of the figure, reveal two clear effects. First, both claiming observed actions as performed and the reverse error of giving away performed actions as observed occur reliably more often than predicted from guessing. Second, the confidence intervals for the two errors overlap considerably. At present, therefore, there is no evidence to suggest that one form of error is more frequent than the other. This argues against mirroring-views of observation inflation, according to which the potential for source-confusion should only exist for observed actions.

My third question was whether a motor system load at encoding would reduce the magnitude of the observation inflation effect. Mirroring views suggest that occupying the motor system would undermine the formation of self-related representations of the partners actions, and therefore reduce source errors. I therefore looked at participants' false recall of their partner's actions depending upon whether they were under any motor load at encoding or not. The summary statistics in Figure 2.7 support two clear conclusions. First,

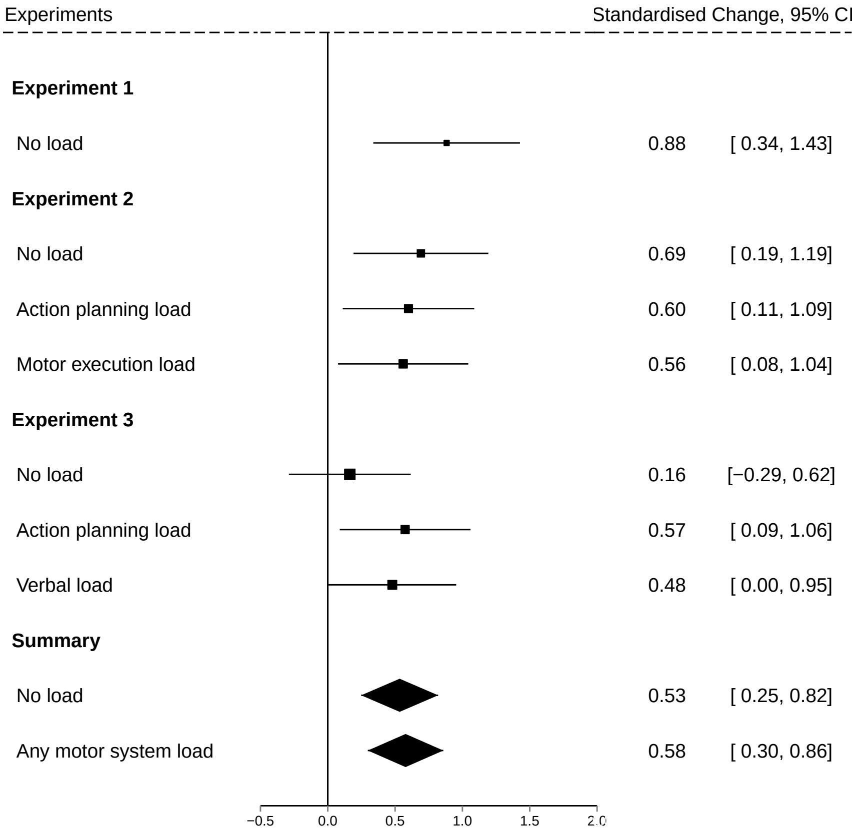

Figure 2.7. Standardized mean change between experimental and predicted source errors committed in the Recall performed task in Experiments 1 through 3. 0 indicates source error rates as would be predicted by guessing. The error bars are 95% confidence interval. The polygons represent the summary effects (fixed effects) across the indicated conditions.

across experiments, observed actions are falsely claimed as performed regardless of concurrent load. Second, the data do not support the predicted reduction in this source error with a motor load: the confidence intervals on the summary statistics of the control and motor load conditions overlap considerably, with the overall pattern suggesting a slight increase in errors with motor load, rather than the predicted decrease.

Finally, Figure 2.7 allows closer comparison of the results in Experiment 3 relative to the prior experiments. The main difference between Experiment 3 and 2 was the change in retrieval task that required participants to carefully examine the actions they were reporting. Rather than examining Experiment 3 in isolation, I can now ask if the change in retrieval task replicated or in fact reduced the memory errors.

I looked at the effect of observed versus predicted source errors with a Bayesian analysis separately for both retrieval tasks following the Dienes (2008) protocol. Here, I can test first whether source errors in Experiment 3 are more compatible with participants guessing (no difference from 0) or more compatible with the source error effect in Experiment 2. Did the extended recall task reduce source errors to guessing if actions were encoded in the No load condition? In other words, the Bayesian approach allows me to specify an exact alternate hypothesis I am testing compared to a generic alternate that is different from the null.

Thus, I first compared the No load condition in both the Recall performed and Recall observed retrieval task from Experiment 3 to the matching condition in Experiment 2. The prior for the Recall performed task was defined by the effect in Experiment 2 with a half-normal distribution, i.e., one-tailed, with $M = 0, SD = 1.28$. The data was defined by the effect in Experiment 3 with a sample mean of $M = .15$ and $SE = .20$. The prior for the Recall partner task was defined by the effect in Experiment 2 with a half-normal distribution, i.e., one-tailed, with $M = 0, SD = 2.04$. The data was defined by the effect in Experiment 3 with a sample mean of $M = .22$ and $SE = .25$. This resulted in a Bayes Factors (BF) of 0.32 for the Recall performed and 0.29 for the Recall observed task. This means that the data in Experiment 3 provide 3 and 4 times more evidence, respectively, in favour of a null memory error effect than in favour of a replication of the effect in Experiment 2. This suggests that careful monitoring reduced memory errors effect when recalling actions observed without a load. This effect is of moderate size.

I next compared the Action planning load condition in the Recall performed task across the two experiments. The prior was defined by the effect in Experiment 2 with a half-normal distribution, i.e., one-tailed, with $M = 0, SD = .95$. The data was defined by the effect in Experiment 3 with a sample mean of $M = .98$ and $SE = .38$. When observed actions were encoded under concurrent Action planning load, the size of the memory error in the Recall performed task in Experiment 3 is more compatible with the size of the effect in the matching condition in Experiment 2 than the null, $BF = 15$. This means that the memory error effect observed in the Action planning condition in Experiment 3 is 15 times more compatible with the memory error effect in Experiment 2 than with a null memory error effect. This suggests that careful monitoring at retrieval can reduce source errors during recall, but only does so when participants were not engaged in a secondary task during the encoding of actions.

2.8 Discussion

2.8.1 Does the observation inflation effect generalise?

The first aim of this series of studies was to test if the observation inflation effect reported by Lindner et al. (2010) generalizes to a different experimental design, retrieval task and with different action stimuli. Across three experiments, I have shown a robust effect of observed actions being falsely retrieved as self-performed actions.

This suggests that the observation inflation effect is not merely an effect of verbalisable action phrases, but can also occur with non-verbalisable, non-object-directed actions. Additionally, it shows that the observation inflation effect observed by Lindner et al. (2010) is not solely a result of the experimental design or retrieval task. In this series of experiments, I observed robust evidence of observed actions leading to false claims of self-performance in two versions of a free recall paradigm.

2.8.2 What is the role of the motor component in source confusion?

In this chapter, I tested whether mirror neuron network activation caused by observing actions is critically responsible for falsely claiming those actions as performed (as predicted by the motor simulation account). I failed to find clear evidence to support this account.

First, an account based on internal replicas of observed actions (Craighero et al., 2002, 1999) does not suffice to explain why self-performed actions were reported as “observed” since motor activation during performance always points towards self-performance, and no conflict between self- and other performance should arise. One might argue that even if the reverse error cannot be explained by mirror neuron activation, observation inflation still can. This argument assumes a) two separate mechanisms to account for very similar errors and b) that observation inflation can in fact be explained by mirror neuron system activation. If mirror neuron activation is fundamentally responsible for later false memories of self-performance, such false memories should be disrupted when observers' motor systems are occupied with a motor task at the time of observation. However, direct manipulations of the extent to which the observed actions could be mirrored at encoding showed no evidence for the expected reduction in subsequent source errors, in either Experiment 2 or 3.

Of course, one possibility is that while my manipulations aimed to disrupt basic motor execution and higher level action planning, it is possible that neither sufficiently or particularly resulted in a reduction in mirror neuron activation. However, concurrent motor action, with very similar procedures, has been shown to impair perception of observed actions (Hamilton et al., 2004; Vetter & Wolpert, 2000; Zwickel et al., 2007), the acquisition of motor skills during imitation and observation learning (Bach, Allami, et al., 2014), and observation of actions has been shown to influence execution of actions (Bach &

Tipper, 2007; Kilner et al., 2003; Press et al., 2011; Zwickel & Prinz, 2012). This suggests that concurrent motor activation should be sufficient to affect mirror neuron network activation during action observation. Indeed, I found that both low-level and high-level motor system load affected the duplication of a partner's actions in the encoding task as well as correct recall of observed actions at test, confirming that our load manipulation was generally effective.

A second possibility is that while the actions I used required motor activation to be performed, that the primary encoding modality was a different one. Participants were asked to perform shapes using their bodies. These were unfamiliar actions requiring them to move their body from a regular standing position into a different action end-state. Rather than drawing on the motor information of pushing their body into a new position, participants may have encoded their own end-state visually. Since they were required to perform a shape, in an effort to make sure that they did in fact perform the shape, it is possible they visualized their own performance from a third-person point of view. In that case, motor information of performance would have not driven responding in a free report task. However, anecdotally, participants did report at retrieval that they drew on "what it felt like" to make a shape.

One way to address this in future research in a variation of this paradigm would be to force the focus on the movement between end-states, by asking participants in a cued-recall paradigm what the second movement was that they moved into after performing a shape.

A third possibility is that I was successful in disrupting the encoding of the motor component of observed actions but additionally disrupted the encoding of other memory components. While the disruption of the encoding of the motor component could be associated with a decrease in source errors, this may have been counteracted by an increase in source errors due to the overall cognitive load at encoding (Craik, 2014). There is some evidence to support this point. Participants showed effects of concurrent load on performed actions, even when those actions were only performed during a concurrent load block rather than under concurrent load. This suggests that the task-switching required, for example, may have driven some of the effects of concurrent load I did observe in the experiment. To address this, Experiment 4 in Chapter 3 implements a manipulation of motor interference at retrieval rather than at encoding.

Finally, Lindner et al. (2016) recently tested the motor activation account of observation inflation directly within their own paradigm. They asked participants to perform motor movements during observation, some compatible with the action they observed at the time, some incompatible. They reported that compatible motor movements increased 'Performed' responses relative to incompatible motor movements. This means that performing the motions of an action increased participants' tendency to later claim they had performed that action at a previous point. Lindner et al. did not find a specific decrease of false 'Performed' responses under incompatible motor activation as would be predicted under a motor activation account.

In sum, source confusion of actions appears impervious to motor interference predicted under a motor activation account even when the actions used are less complex. This could be because action memories are rich in other qualitative features that are preferentially used for source judgements or because the motor interference tasks used across the experiments did not task the right aspects of the motor system at the right time.

2.8.3 Can the source monitoring framework account for the data?

As I suggested earlier, and has been discussed by Lindner et al. (2010), a clear alternative to a motor activation account would be the source monitoring framework (Johnson et al., 1993). Here, the source of any given memory is inferred from qualitative features encoded alongside the item information when people recall a memory. Source confusion arises when those qualitative features are not available and/or if participants neglect to evaluate them. I propose that the source errors observed here and in the observation inflation paradigm can be very well accounted for in this framework.

A source monitoring account can explain all features of the data in our experiments and has been previously suggested to account for false memories of self-performance after visualization of actions (Henkel & Carbuto, 2008; Lindner & Henkel, 2015). Firstly, a source monitoring account can explain why participants not only misremember observed actions as performed but also performed actions as observed. Performed and observed actions both create memories of events that share similarities: both include not only proprioceptive or motoric features such as actions' trajectories or the manner of performancem but also visual information such as whether someone else's face was in view and whether the body parts belonged to this person. These commonalities predict not only source errors in the Recall performed task, but a general confusion about the origin of encoded actions that would affect both tasks equally, and more so if these disambiguating aspects are not encoded. Recall, in Lindner et al. (2010)'s paradigm participants first encoded actions by reading a verbal description or performing the action. They subsequently watched a video clip of someone performing the action. It seems likely that observing actions in Lindner et al. (2010) paradigm led to rich encoding of actions in the observation condition, with resulting memory events containing features more diagnostic of 'performed' than 'read' action phrases. Since participants were never asked to decide between actions having been 'Performed' or 'Observed', 'Performed' responses would represent the next likely response alternative. Schain et al. (2012) showed that source cues can reduce observation inflation, and Lindner et al. (2012) argued that observing an in-group member increases observation inflation. Both effects are compatible with the source monitoring framework.

Secondly, a source monitoring account predicts that source errors decrease when participants evaluate all qualitative features of individual memories systematically. Lindner et al. (2010, Exp. 2) manipulated retrieval task instructions in two ways. Some participants were given a warning about the observation inflation effect, while others were asked to consider the qualitative features of each memory item as they made their recognition decision. Neither resulted in a decrease in observation inflation, leading Lindner et al. to reject a source monitoring account on the basis of these null effects. In contrast, the retrieval task instructions I used in Experiment 3 (asking participants to make explicit source decisions) did eliminate source errors. This suggests that allowing participants to withhold remembered items from report is an effective strategy of decreasing the reports of source errors when retrieving action events by preventing participants from neglecting to consider the source of recalled memories. This is in line with effects that showed that source performance can be drastically affected by the instruction and decision required at test (Dodson & Johnson, 1993; Marsh & Hicks, 1998; Marsh et al., 1997). Interestingly, finally, that same retrieval task manipulation failed to show the same elimination of observation inflation when actions had been encoded under concurrent load. Under source monitoring account predictions, any distraction that prevents the encoding of qualitative features of items should result in a poorer source memory trace, as fewer source-relevant features might be encoded. Our data suggest that concurrent load prevented the encoding of source-diagnostic information for individual actions to such a degree that even systematic consideration of source at test was unable to overcome that impairment.

Chapter 3

Motor component - retrieval manipulation

3.1 Background

Observation of actions leads to false memories of self-performance at a subsequent retrieval test. In Chapter 2, I showed that this observation inflation effect discussed by Lindner et al. (2010) using the observation inflation paradigm with recognition test is also observed using a recall test. When recalling actions they performed themselves, participants sometimes include actions they observed the other person perform. Lindner et al. claimed that the crucial mechanism responsible for these false memories is the creation of motor representations in the observer's motor system at the time of observation (the motor simulation account). This motor representation is reactivated at retrieval and leads participants to misremember observed actions as self-performed. There are three possible approaches to critique and test this account.

First, this account does not predict that participants would falsely retrieve self-performed actions as observed. In fact, I showed throughout the experiments in Chapter 2 that participants commit this reverse error when given the chance.

Second, if the motor representation cannot be encoded during observation, fewer observed actions should be falsely retrieved as self-performed. For example, in line with selective interference effects shown by Saltz and Donnenwerth-Nolan (1981), I asked participants to perform motor tasks during observation in Experiments 2 and 3. This interference, if anything, increased the number of errors participants made during encoding. This is not compatible with the motor simulation account. However, it is possible that the concurrent motor interference condition I used did not precisely (or modularly) affect the motor trace.

The aim of the concurrent motor manipulation was to disrupt the motoric processing of observed actions, ensuring that when memory of those events was consolidated, the memory events would lack motor information. It is possible that the concurrent load in

Experiment 2 and 3 did not interfere with encoding of observed actions that selectively. Attending to the secondary task may have interfered with processing the conceptual or visual aspects of the observed actions, or more broadly interfered with the consolidation and rehearsal of those memories. In fact, the data showed that concurrent load at encoding affected the retrieval of performed actions in similar patterns as it affected the retrieval of observed actions, even though performed actions were never directly encoded under concurrent load. This pattern of data is only plausible if the mere presence of a secondary task provided enough distraction that it prevented consolidation or rehearsal of any actions (performed or observed) processed during these concurrent load blocks. This is compatible with the view that any distraction should disrupt correct source responding (Johnson et al., 1993; Macrae et al., 1999), but does not provide clear evidence against a specific motor simulation account of observation inflation.

The third approach to test the motor simulation account is therefore to interfere with the retrieval of the motor trace when participants are asked to make their source judgements, after they encoded actions under full attention. If observed actions are falsely retrieved as self-performed because of the encoded motor representation, preventing participants from accessing this trace should reduce the number of errors.

Crucially, the motor simulation account assumes that motor information is not only encoded during performance or observation, but that these components of the memory trace are accessed when making retrieval decisions. There is some evidence that actions that were previously performed lead to activation of motor areas at retrieval. Masumoto et al. (2006) and Nyberg et al. (2001) asked participants to encode items verbally or by enactment. They showed that when participants retrieved actions they had performed previously, they showed activation (measured by MEG and PET respectively) in the primary motor cortex and left parietal cortex for those actions, but did not show that activation for items they had only encoded verbally. In line with those results, Brandt et al. (2013) observed participants showing activation in the supplementary motor area and ventrolateral PFC when participants retrieved actions they had performed or imagined. Activation of motor areas is not limited to the retrieval of self-performed actions. During encoding in Senkfor et al. (2002), participants performed actions with objects, watched the experimenter perform actions with objects, imagined performing actions with objects or estimated the cost of objects involved in actions. At test, participants were shown photos of objects and asked to make Old/New recognition judgments. ERP recordings revealed activation in fronto-central areas roughly indicating the premotor cortex areas during retrieval of action memories (performed, watched and imagined conditions) but not during retrieval of cost estimation memories. While activation was quantitatively higher for performed actions over watched and imagined actions, there was no qualitative difference in the activation. Senkfor (2008) showed that the effects are similar if actions are pantomimed rather than acted out with objects, though activation after performance with objects was stronger.

If motor areas are activated during the retrieval of performed actions to make recognition judgements, arguably, there should be a benefit of re-enacting actions at retrieval. Indeed, Engelkamp et al. (1994) showed that recognition of motor actions was benefited by re-enactment, but that critically this was only the case if participants used the same hand to perform the action at encoding and retrieval. While Mulligan and Hornstein (2003) replicated the re-enactment effect overall, they could not replicate the specific handed-ness effect. Importantly, Engelkamp et al. asked participants to encode and retrieve action verbs, while Mulligan and Hornstein asked participants to retrieve verb-object phrases. This means that recognition judgements in Engelkamp et al. relied more strongly on motor movement (alongside semantic and conceptual information) and increased the importance of the exact replication of the encoding movement at retrieval. On the other hand, recognition judgements in Mulligan and Hornstein could be made on the basis of object-cues without considering motor information at all. The effects of re-enactment for their study may therefore be due to general context reinstatement at retrieval rather than reinstatement of the exact motor trace.

There is sufficient evidence, in sum, to suggest that motor activation from encoded actions is reactivated at retrieval. If that is the case, preventing the reactivation of the motor trace at retrieval should reduce confusion over the source of observed actions. Experiment 4 will test whether selective interference at retrieval can reduce the rates of false performed responses by increasing participants' ability to make correct source decisions about observed actions.

3.2 Reasoning for the methodology

I will employ a different paradigm in Experiment 4 compared to the previous experiments, while the theoretical question remains largely the same. The changes and the reasoning behind them are as follows.

First, participants will perform part of the retrieval test under interference conditions. Participants encode actions without interference. At test, I will ask participants to perform simple arm movements during half of the recognition test to prevent reactivation of motor traces at retrieval. To ensure that any observed effects are not simply the result of interference in general, half of the participants are given a visual interference task at retrieval, rather than the motor interference task.

Second, I will ask participants to generate actions that are performed with objects rather than ones that are performed only with the use of their bodies as in Experiments 1 through 3. The actions in those previous experiments necessitated that participants re-performed the actions at recall since verbal recall was not possible. This would make manipulating the retrieval of the motor trace with concurrent motor interference at retrieval impossible, since participants would have to perform secondary movements while re-performing the actions from encoding. Therefore, I will ask participants to generate actions by using objects from a grid of objects to make them interact (similar to R. L. Cohen, 1983). Participants will then complete a verbal source recognition test at retrieval. One concern with verb-object phrases is that participants could make recognition decisions on the basis of the object alone (compare Engelkamp et al., 1994; Mulligan & Hornstein, 2003). Two aspects of the paradigm should limit this: a) participants will use the same subsets of objects for various items, so objects were not diagnostic of source, and b) the distractors at test will be matched to the verb-object phrases participants generate by using the same objects and exchanging the motor action. If a participant generated ‘Put the card in the mug’, a possible distractor could be ‘Push the mug with the card’. Participants now have to take the motor action into account to make correct recognition judgements.

Third, I will change the retrieval task. Experiments 1 through 3 extended the observation inflation effect to a recall paradigm. I additionally showed that participants commit the reverse error when given the opportunity and falsely recall self-performed actions as observed. It is possible that this reverse error is only an artefact of the free recall task I used. When participants recall observed actions, they have to search their memories for the weaker source of items and may simply report (the stronger memories of) performed actions because they happen to come to mind, in line with the mistaken fluency account of source errors (Jacoby, Woloshyn, & Kelley, 1989). For this reason, I will use a 3AFC (Performed, Observed, New) recognition paradigm in Experiment 4. Rather than adding observation as an additional manipulation as in observation inflation paradigms

(Lindner & Davidson, 2014; Lindner & Echterhoff, 2015; Lindner et al., 2010; Lindner & Henkel, 2015; Lindner et al., 2016, 2012; Schain et al., 2012), I will contrast performance and observation at encoding as in other studies of source recognition of actions (e.g. Hornstein & Mulligan, 2004; Leynes & Kakadia, 2013; Rosa & Gutchess, 2011). In line with previous research on source memory of actions, I expect that participants will claim observed actions and self-performed and self-performed actions as observed. To avoid ceiling effects in recognition performance, the delay between encoding and retrieval will be set to a week compared to the day's delay in Experiments 1 through 3.

3.3 Analytic approach

I will first analyse the data using the approach typically used for 3AFC source recognition data (e.g., Foley, Bays, Foy, & Woodfield, 2015). Here, participants' item memory and source discrimination is assessed separately. Participants' item memory is defined as their ability to distinguish encoded items from novel items, i.e., make Old/New decisions. This results in hit rates and false alarm rates for item memory performance. Since participants may have different response tendencies or biases (i.e., be more likely to respond 'Old' overall), I will report a measure that corrects for these biases d', which contrasts the norm scores of hit rates and false alarm rates.

I will analyse participants' source memory performance using the conditional source identification measure (CSIM), also known as the Identification-of-Origin measure (IDO). CSIM gives the rate with which correct source judgements are made once old items were correctly identified as old (e.g. Foley et al., 2015; Hornstein & Mulligan, 2004). Additionally I will report d' to estimate participants' source discrimination independent of source response biases and c as a measure of response bias in participants' source discrimination.

While the CSIM appears to give an unbiased, independent estimation of source discrimination at test, Murnane and Bayen (1996) showed that it is contingent on item recognition and is subject to influences of response bias or guessing. While this is less concerning when item recognition is equal across item types, it makes CSIM difficult to interpret when item recognition differs as a function of encoding condition (see also DeCarlo, 2003).

Murnane and Bayen (1996) suggested fitting multinomial processing models to address this issue (also see Batchelder & Riefer, 1990). Fitting these models allows separate estimation of item parameters, source parameters and guessing parameters. As I discussed in the Introduction, multinomial processing models make the assumption of rectangular memory distributions underlying performance in a memory test. This is under debate (Batchelder & Alexander, 2013; Pazzaglia et al., 2013; Slotnick & Dodson, 2005; Yonelinas & Parks, 2007). In addition to fitting a multinomial processing model (more specifically, the Two-High-Threshold model (Bayen et al., 1996)), I will therefore also fit signal detection models. Those models assume that memory distributions are best represented by Gaussian distributions.

There are two types of signal detection models that could account for source monitoring performance: the two-dimensional signal detection model which formalises the source monitoring framework (DeCarlo, 2003; Hautus et al., 2008; Johnson et al., 1993; Slotnick & Dodson, 2005) and the one-dimensional signal detection model that formalizes responding based on the relative strength of sources (Marsh & Bower, 1993). I will fit both types of model to test which better accounts for the data.

For the results of this experiment, I will therefore first report the model-free analysis of item memory performance and source discrimination, with an additional section on inflation measures. I will discuss those results, then report model-based analyses that contain model fits of Two-High-Threshold, two-dimensional and one-dimensional signal detection models. In the final section, I will compare the fits of those three models to comment on which better represents participants' performance in this experiment.

3.4 Aims and predictions

In Chapter 2, I explored a motor simulation account of false retrieval of observed actions as 'Performed' using two different approaches: testing a) if performed actions are also falsely identified as 'Observed', and b) if interference of motor encoding during performance prevents false memory retrieval. Under a motor simulation account, it is the retrieval of the motor trace that leads to false memories of having performed an action after having observed it. In Chapter 3, I therefore explore if motor interference at retrieval prevents false retrieval of observed actions as performed by preventing reactivation of motor traces.

Experiment 4 employs a different experimental paradigm to explore the same theoretical question as the experiments in Chapter 2. In this experiment, participants will take turns performing actions using objects. At test, participants will complete a 3AFC (self-performed own actions, observed partner action, novel actions) source recognition test, with some of the retrieval completed under interference. In addition to testing the effect of motor interference at retrieval, I will also test if visual interference affects performance.

As in Experiments 1 through 3, I will first look at participants' item memory. In line with the enactment effect (e.g. Engelkamp et al., 1994), I expect that participants' item recognition of performed actions will be higher than their recognition of observed actions. I do not predict an effect of motor or visual interference on item recognition.

Second, I will test if participants commit source errors, that is falsely identify observed actions as 'Performed' and performed actions as 'Observed'. In this 3AFC source recognition paradigm, this should be reflected in CSIMs for performed and observed actions being below ceiling. Following that, I will look at source errors specifically with the inflation metric proposed by Lindner et al. (2010) to test if encoding actions by performance or observation increases the rates of false responding relative to false responses participants make to novel items.

The main predictions of this experiment concern the effect of interference on participants' ability to discriminate performed from observed actions. Under the assumption of the diagnostic features account (Leynes & Kakadia, 2013), performed and observed actions are characterized by certain features, with, for example, the motor trace typically associated with performed and not with observed actions. Under a motor simulation account, participants create memories of actions by observing others perform actions. Those memories contain semantic, conceptual and visual information, but are additionally enriched by motor representations created in the observer's own motor system. Similarly, participants encode actions they perform themselves primarily motorically, with additional encoding of semantic, conceptual and visual information. For observed actions, the visual memory trace of an observed action clearly points towards having observed the action (seeing another person act out the action), while the motor trace indicates having performed this action oneself. This conflict increases false responding.

Under this account, when participants are prevented from accessing the motor trace of actions during a recognition test by concurrent motor interference, they will have to rely on the non-motoric components of the memory trace to make their decisions. This should prevent the false retrieval of observed actions as self-performed, since judgements about those actions are now based largely on visual information that should indicate having observed someone else perform it. The complementary prediction is that this may decrease the rate with which self-performed actions are correctly identified as self-performed. When attempting to identify self-performed actions participants cannot rely on the motor cue, and may be more likely to misidentify a performed action as observed.

Under a source monitoring account, any interference at retrieval may limit source memory performance (e.g. Johnson et al., 1993). If motor traces are not encoded for observed actions, motor interference would lead to participants' being less able to identify performed actions, with no benefit to the identification of observed actions.

In addition to motor interference, I will also test participants' source discrimination under visual interference. Visual interference should increase false identification of observed actions as self-performed. Without the visual cue, participants have to rely largely on the misleading motor cue to make their decisions, leading them to misattribute the action. While visual information of self-performance is indicative of having performed the action oneself, participants should still be able to make us of the motor information in absence of this cue. The source monitoring account would predict a similar result here.

3.5 Experiment 4

3.5.1 Method

Participants

52 members of the public participated for payment of £8. 6 participants did not attend all sessions and their data were excluded.

Stimuli

36 objects were split into 6 sets of 6 objects each (see Appendix B.1). Participants generated their own verb-object phrases by choosing two objects from a set at a time to make them interact with one another, resulting in phrases like ‘Push the mug with the pen knife’. Participants were instructed to use a variety of combinations of objects throughout the trials for one set of objects to ensure that both participants would interact with the majority of objects in a set.

Procedure

The experiment consisted of two sessions that took place a week apart. Each session lasted approximately 30 minutes. Participants attended the first session in pairs. They were seated to face one another across a table. Across the encoding phase, 6 sets of objects in total were placed on the table between them, one set at a time. Participants were asked to take turns performing actions by always using two of the objects at a time to make them interact. Participants generated the ideas for the action phrase they performed (that is which objects they made interact in which way) themselves.[1] Participants did not verbalize the actions they performed, they only used the objects to enact them. Once participants had finished performing the action they had generated, participants returned the objects to their original position on the table. When it was their partner's turn to perform an action phrase, participants observed their partner perform the action. Each participant generated 6 actions for each of the 6 sets of objects. Participants were encouraged to use a variety of objects and a variety of motor actions across a set. They were instructed to avoid repeating actions (that is, the exact combination of objects and motor action) that had already been performed by them or their partner. The experimenter verbally coded and noted down the actions performed as action phrases.

Participants returned individually a week later to complete a 3AFC (Own, Partner, New) source recognition test. The test list consisted of the 6 actions participants had performed for each set, 6 actions they had observed their partner perform and 6 distractor items that were new at test for a total of 36 performed, 36 observed and 36 novel actions at test. The actions were read to participants by the experimenter. Participants responded verbally whether they had performed the action themselves by saying 'Own', had observed their partner perform the action by saying 'Partner' or if the action had not been performed by them or their partner at encoding by saying 'New'.

Distractor actions were generated individually for each pair of participants. These actions were created by re-using combinations of objects these particular participants had used at encoding. 3 new actions were matched to objects used by one, the other 3 to match objects used by the other participant. The action verbs were equally matched to action verbs used by participants when they generated the actions at encoding. For correct discrimination of old from novel actions, participants were now required to retrieve the exact action performed at encoding.

At the beginning of the test phase, participants were randomly assigned to one of two groups: the Visual interference group (Visual group) or the Motor interference group (Motor group). They completed the source test on a set-by-set basis to allow within-

---

[1] I will refer to these action items as actions. Each action contains 2 objects and a motor action that connects the two objects (a verb). Action therefore refers to the complex of 2-objects-and-verb rather than only the verb or motor action itself.

participant manipulation of interference at retrieval. In both groups, participants completed the recognition test for half the sets without interference (the control condition), and for the other half with interference (the interference condition). The assignment of sets to control or interference conditions as well as the order of control and interference condition was counterbalanced across participants.

Visual interference consisted of a continuous sequence of faces on a laptop screen. Each face was shown for half a second without an interstimulus interval. The stream of faces was created using the funelled faces from the Labeled Faces In the Wild database (Huang, Jain, & Learned-Miller, 2007; Huang, Ramesh, Berg, & Learned-Miller, 2007). Each photo used in the stream measured 250 by 250 pixels and showed a face from the shoulders up in a range of real-world contexts. The faces were presented in the middle of the laptop screen. Participants were instructed to keep looking at the faces as they performed the retrieval task, and corrected to return to look at the faces if they looked away during the experiment. Participants looked at the faces continuously throughout this part of the retrieval task, i.e., as actions were read to them, as they responded and in-between items.

For the interference condition in the Motor group, participants were asked to close their eyes and move their arms in a pre-set ('Macarena dance'-like) pattern. Participants practiced the movement prior to the test phase. They were reminded to continue to move their arms throughout, when actions were read to them, when they were responding or in-between items.

3.5.2 Model-free analysis

Encoding phase

I first looked at the errors participants committed during the encoding phase. Even though participants were instructed not to, they occasionally repeated actions they or their partner had previously generated. Encoding phase errors were very rare. 2 out of 52 participants repeated actions they had performed themselves already and 8 out 52 repeated actions their partner had performed. Expressed as an average percentage of performed actions, self-repeats occurred in 0.11% (SD=0.01%) of performed actions and partner-repeats occurred in 0.53% (SD=0.01%) of performed actions. As in prior experiments, partner-repeats were excluded from the subsequent analyses to ensure that each action was uniquely performed by one person in the pair.

Table 3.1. Responses in Experiment 4

	Visual group			Motor group		
	Performed	Observed	Novel	Performed	Observed	Novel
Control						
'Own'	.62 (.14)	.18 (.10)	.14 (.09)	.54 (.19)	.21 (.10)	.20 (.10)
'Partner'	.22 (.10)	.53 (.14)	.29 (.13)	.23 (.13)	.48 (.11)	.25 (.13)
'New'	.16 (.09)	.29 (.10)	.58 (.16)	.23 (.16)	.30 (.14)	.55 (.18)
Interference						
'Own'	.63 (.13)	.15 (.12)	.15 (.09)	.61 (.18)	.19 (.13)	.15 (.10)
'Partner'	.22 (.12)	.57 (.17)	.28 (.17)	.23 (.14)	.50 (.13)	.24 (.15)
'New'	.15 (.11)	.28 (.12)	.57 (.20)	.16 (.10)	.31 (.17)	.61 (.18)

Note. Average performance by item (Performed, Observed and Novel in the columns) and response type ('Own','Partner' and 'New'). Proportions of responses for each item type sum to 1, in both groups and interference conditions. Data is based on 23 participants in each group and on average 18 items per condition and item type.

Test phase

Given the complexity of the experimental design and retrieval task, the resulting data can be analysed in various ways. I tested the effect of visual interference separately, by creating a Visual group and a Motor group, each containing a control condition in addition to the interference condition. Since my predictions were focused on the impact of interference relative to control, I will report results for both groups separately. Table 3.1 shows the proportion of responses by type of item for all groups and conditions.

I will first present results from a conventional, model-free analysis, beginning with participants' item memory and moving on to participants' source discrimination. This analysis is typically used in 3AFC source recognition tasks (e.g., Foley et al., 2015). Following that, I will look at source errors specifically with the inflation metric proposed by Lindner et al. (2010).

Item Memory I first looked at participants' item memory, i.e., participants' ability to distinguish old actions (that they or their partner performed at encoding) from distractor actions that were novel at test. I did not predict any effect of the interference manipulation on item memory performance. I did predict that in line with the enactment effect, item memory for performed actions should be higher than for observed actions (e.g. Engelkamp et al., 1994).

I followed the analysis used by Hornstein and Mulligan (2004) in a similar experimental design. Hits indicate participants responding 'Own' or 'Partner' (rather than 'New') to either performed or observed actions. The hit rates for performed actions indicate the proportion of performed actions correctly classified as old (i.e., participants responded 'Own' or 'Partner'). Hit rates for observed actions represent the proportion of observed actions correctly recognised as old. False alarms are similarly identified as the proportion of novel actions that participants falsely classified as old. Hit rates and false alarm rates are given in the top of Table 3.2.

I analysed participants' hit rates separately for the Visual and Motor groups with 2 Source (Performed, Observed) x 2 Interference condition (Control, Interference) repeated measures ANOVAs.

In the Visual group, there was a significant main effect of Source, $F(1,22) = 51.69, MSe = 0.01, p < .001, \eta_p^2 = .701$. Performed actions were correctly recognized as old more frequently than observed actions (the enactment effect). There was no significant main effect of interference condition or significant interaction, both $F < 1$.

In the Motor group there was a significant main effect of Source, $F(1,22) = 23.95, MSe = 0.01, p < .001, \eta_p^2 = .521$. Performed actions were correctly recognized as old more frequently than observed actions. There was no significant main effect of interference condition, $F(1,22) = 2.62, MSe = 0.06, p = .12, \eta_p^2 = .106$, but there was a marginally significant interaction, $F(1,22) = 4.27, MSe = 0.01, p = .051, \eta_p^2 = .163$. Looking at the data separately by source shows a significant main effect of interference condition for performed actions, $F(1,22) = 6.46, MSe = 0.01, p = .019, \eta_p^2 = .227$. Under motor interference, participants were more likely to correctly recognize performed actions as 'old'. The same effect was not significant for observed actions, $F < 1$.

I next looked at false alarms, that is actions that were novel at retrieval, that participants judged to be old (i.e., 'own' or 'partner'). In the Visual group, there was no significant evidence for an effect of interference on the frequency of false alarms , $F < 1$. That pattern was similar in the Motor group, $F(1,22) = 2.74, MSe = 0.01, p = .11, \eta_p^2 = .111$.

Both hit rates and false alarm rates may be influenced by response bias and therefore not give an unbiased estimate of participants' item memory. I decided to explicitly address the response bias concern and calculated d' for each individual and condition. d' gives an indication of the discriminability of two types of items, here: performed and novel actions as well as observed and novel actions, while correcting for the influence of response bias by taking both hit rates and false alarm rates into account. I followed the procedures suggested in Macmillan and Creelman (2004) and calculated d' as

Table 3.2. Item memory in Experiment 4

	Visual group		Motor group	
Actions	Control	Interference	Control	Interference
Hits				
Performed	0.84 (0.09)	0.85 (0.11)	0.77 (0.16)	0.84 (0.10)
Observed	0.71 (0.10)	0.72 (0.12)	0.70 (0.14)	0.69 (0.17)
False Alarms	0.42 (0.16)	0.43 (0.20)	0.45 (0.18)	0.39 (0.18)
d'				
Performed	1.34 (0.74)	1.40 (0.80)	1.04 (0.81)	1.45 (0.64)
Observed	0.79 (0.62)	0.80 (0.62)	0.77 (0.72)	0.85 (0.74)

Note. Means with standard deviations in brackets. d' indicates discrimination of old and novel actions, free from response bias.

$$d' = z(H) - z(FA) \tag{3.1}$$

where normalized scores of Hit rate (H) and False Alarm rate (FA) are contrasted. This assumes that items are drawn from an underlying normal distribution with equal variance for item types. d' can vary between 0 and infinity, with infinity occurring with hit rates of 1 and false alarm rates of 0. In case of those values, hit rates and false alarm rates are adjusted by $1 - \frac{1}{2N}$ and $\frac{1}{2N}$ respectively where N is the number of items the calculation is based on (here: the number of performed, observed and novel actions respectively).

I tested participants' item discrimination d' with a 2 Source (Performed, Observed) x 2 Interference condition (Control, Interference) repeated measures ANOVAs separately for the Visual and Motor group.

In the Visual group, there was a significant main effect of Source, $F(1,22) = 29.11$, $MSe = 0.27$, $p < .001$, $\eta^2_{\text{p}} = .570$. Discrimination between old and new actions was higher for performed than for observed actions, the enactment effect. There was no significant main effect of Interference condition, nor a significant interaction, both $F < 1$.

In the Motor group, there was a significant main effect of Source, $F(1,22) = 19.54, MSe = 0.22, p < .001, \eta^2_{\text{p}} = .470$. Discrimination was higher for performed than observed actions. There was a marginally significant effect of Interference condition, $F(1,22) = 3.20, MSe = 0.41, p = .088, \eta^2_{\text{p}} = .127$ and a marginally significant interaction, $F(1,22) = 4.27, MSe = 0.15, p = .052, \eta^2_{\text{p}} = .161$. I looked at the effect of interference separately for performed and observed actions. There was a significant effect of Interference condition for performed actions, $F(1,22) = 9.09, MSe = 0.21, p = .006, \eta^2_{\text{p}} = .292$. Discrimination of performed from novel action was more successful when participants were making their decisions under

Table 3.3. Responses to novel items in Experiment 4

	Visual group		Motor group	
	Control	Interference	Control	Interference
‘Own’ responses				
Distractors matched to performed actions	0.13 (0.15)	0.17 (0.13)	0.19 (0.13)	0.14 (0.13)
Distractors matched to observed actions	0.15 (0.12)	0.14 (0.09)	0.20 (0.13)	0.16 (0.13)
‘Partner’ responses				
Distractors matched to performed actions	0.28 (0.17)	0.27 (0.19)	0.25 (0.19)	0.25 (0.17)
Distractors matched to observed actions	0.29 (0.20)	0.29 (0.24)	0.25 (0.16)	0.23 (0.18)

Note. Means with standard deviations in brackets. The means give the proportion of distractors matched to an item type that participants classify as ‘Own’ or ‘Partner’.

motor interference. This effect was not predicted. There was no significant effect when discriminating observed from novel actions, $F < 1$.

Finally, I looked at biases participants showed in responding to novel actions. The distractors at test were generated to match combinations of objects used by the pair of participants at encoding. Half the distractors were matched to self-performed actions, half were matched to observed actions. I tested whether responding ‘Own’ or ‘Partner’ differed by Match. Did combinations of objects, rather than actions, inform participants’ responding? The means are given in Table 3.3.

I analysed the data in 2 Response (‘Own’, ‘Partner’) x 2 Match (To performed actions, To observed actions) x 2 Interference condition (Control, Interference) repeated measures ANOVAs separately for the Visual and Motor group.

In the Visual group, there was a significant main effect of Response, $F(1,22) = 18.24, MSe = 0.05, p < .001, \eta_{\text{p}}^2 = .480$. Responding ‘Partner’ occurred more frequently in response to novel actions than responding ‘Own’. There were no other significant main effects or interactions, $F < 1$.

In the Motor group, there was similarly a significant, though smaller, main effect of Response, $F(1,22) = 6.27, MSe = 0.04, p < .001, \eta_{\text{p}}^2 = .222$. There was no significant main effect of Interference condition, $F(1,22) = 2.74, MSe = 0.01, p = .11, \eta_{\text{p}}^2 = .111$. The remaining main effects and interactions were not significant, all $F < 1$.

This means that participants overall were more likely to classify novel actions as ‘Partner’ than ‘Own’ actions, the ‘It had to be you’ effect (Johnson et al., 1993). There was no evidence that the match of objects to performed and observed actions or the interference

Table 3.4. Identification of origin scores in Experiment 4

	Visual interference		Motor interference	
	Control	Interference	Control	Interference
Source Hits				
Performed actions	0.73 (0.13)	0.74 (0.13)	0.69 (0.18)	0.72 (0.18)
Observed actions	0.75 (0.15)	0.79 (0.18)	0.70 (0.13)	0.74 (0.16)
Source discrimination				
d'	1.40 (0.73)	1.77 (1.24)	1.16 (0.83)	1.43 (1.00)
c	0.03 (0.26)	0.11 (0.34)	0.02 (0.33)	0.03 (0.30)

Note. Calculations are conditional on actions correctly identified as old (‘Own’ or ‘Partner’), SDs in brackets. d'

condition otherwise affected the rates of false responding to novel items.

Source Memory I looked at participants’ source memory performance using the conditional source identification measure (CSIM), where source discrimination conditional on item identification is analysed. The rates with which items correctly identified as old were identified correctly as ‘Performed’ and ‘Observed’ are shown in the top rows of Table 3.4. Under a motor simulation account, motor interference at retrieval should interfere with the retrieval of items’ motor trace. This should result in an increase in participants’ ability to correctly identify observed actions and a decrease in their ability to correctly identify performed actions (i.e., a shift in response bias). Visual interference, on the other hand, should interfere with participants’ ability to retrieve items’ visual trace. This should result in a decrease in participants’ ability to correctly identify observed actions.

I analysed the proportions of actions correctly identified as old that were subsequently correctly identified by source with a 2 Source (Performed, Observed) x 2 Interference condition (Control, Interference) ANOVA with repeated measures on the second factor separately for the Visual and Motor group.

In the Visual group, there were no significant effects of Source or Interference condition, nor was the interaction significant, all $F < 1$. The pattern was similar in the Motor group, with no significant effects of Source, $F < 1$, or Interference condition, $F(1,22) = 1.50, MSe = 0.02, p = .23, \eta_p^2 = .064$, nor was there a significant interaction, $F < 1$.

As with the item memory analysis above, I also looked at source discrimination and response bias with calculation of d' and c since I predicted a shift in response bias by interference. I compared discrimination of performed from observed actions directly, in line with Dodson and Schacter (2001). d' was calculated as before. c was calculated as

$$c = \frac{-(z(H_O) + z(F_O))}{2} \tag{3.2}$$

where H_O indicates the hit rate for performed actions and F_O the false alarm rate for performed actions. This measure indicates how biased participants' responding in the source test is. A positive value indicates participants are biased to respond 'Partner' while a negative value indicates that participants are biased to respond 'Own', with values around 0 suggesting unbiased responding.

The predictions for these measures are more complicated, since they combine the predicted effects on both sources. Under motor interference, better identification of observed and worse identification of performed actions means that the two effects should tend to cancel one another out in the source discrimination measure when comparing a change from control to interference condition. On the other hand, the motor simulation account predicts that participants are less likely to respond 'Own' in the motor interference than in the control condition overall, i.e., a shift in response bias. Under visual interference, source discrimination should be worse under interference.

I looked at source discrimination and response bias separately in the Visual and the Motor group. The means are shown in the bottom of Table 3.4. There was no significant evidence that source discrimination differed between control and interference condition in the Visual group, $F(1,22) = 2.12, MSe = 0.72, p = .15, \eta^2_{\text{p}} = .091$, or in the Motor group, $F(1,22) = 1.55, MSe = 0.52, p = .23, \eta^2_{\text{p}} = .066$. Across all groups and conditions, there was no significant evidence for the response bias c differing from 0, $t < 1.61, p > .12$, i.e., there was no evidence that participants were biased in their source responding. Additionally, there was no evidence that response bias differed between Interference conditions in the Visual group, $F(1,22) = 1.63, MSe = 0.05, p = .22, \eta^2_{\text{p}} = .069$, or in the Motor group, $F < 1$.

In summary, I did not observe the predicted effects of Interference condition in either Visual or Motor group in the source memory measures, regardless of looking only at the proportion correctly identified or the measures that more precisely examined discrimination and response bias.

Inflation rates Finally, I looked at inflation rates to compare this experiment to Lindner et al. (2010) and estimate if encoding actions in this experiment inflates false source responding at test relative to baseline. Recall that Lindner et al. asked participants to per-

Table 3.5. False source responses in Experiment 4

	Visual group		Motor group	
	Control	Interference	Control	Interference
‘Own’ responses				
To observed items	0.18 (0.10)	0.15 (0.12)	0.21 (0.10)	0.19 (0.13)
To novel items	0.14 (0.09)	0.15 (0.09)	0.20 (0.10)	0.15 (0.10)
‘Partner’ responses				
To performed items	0.22 (0.10)	0.22 (0.12)	0.23 (0.13)	0.23 (0.14)
To new items	0.29 (0.13)	0.28 (0.17)	0.25 (0.13)	0.24 (0.15)

Note. Mean proportion of false source responses (SDs in brackets) in response to old actions and novel actions.

form or read action phrases. In a second encoding phase, participants observed someone else perform action phrases, some that had been performed by participants, some read, and some were new. Lindner et al. contrasted participants’ tendency to falsely respond ‘Performed’ to actions after observation (that is, respond ‘Performed’ to read action phrases and novel actions that were observed) with participants’ tendency to falsely respond ‘Performed’ to actions that had not been observed (read actions and actions that were novel at test). In a typical observation inflation study this means taking the difference of a proportion out of a 10 total items and a proportion out of a 25 possible items. The inflation effect now constitutes the extent with which false responses after observation exceed false responses to items that were not observed.

In the present experiment, observation constituted the only encoding of actions. I can contrast false ‘Own’ responses to observed actions with false ‘Own’ responses to actions that were novel at test, in approximation of the inflation calculation used by Lindner et al. (2010) that accounts for an overall guessing of responses. In this case, this means taking the difference of a proportion out of 18 total items (observed actions) and a proportion out of a further 18 possible items (novel actions at test). This assumes that base rates are equally likely for both sets of items. For inflation effects, I expect to see false ‘Own’ or ‘Partner’ responses to old actions exceeding the rates of ‘Own’ or ‘Partner’ responses to novel actions. The inflation effects are presented in Figure 3.1.

For statistical comparison, Lindner et al. (2010) estimated observation inflation by contrasting false ‘performed’ responses to observed actions with false ‘performed’ responses to novel actions. To match their analysis, I performed separate 2 Response type (To old items, To new items) x 2 Interference condition (Control, Interference) repeated measures ANOVAs separately for false ‘Own’ responses and false ‘Partner’ responses and

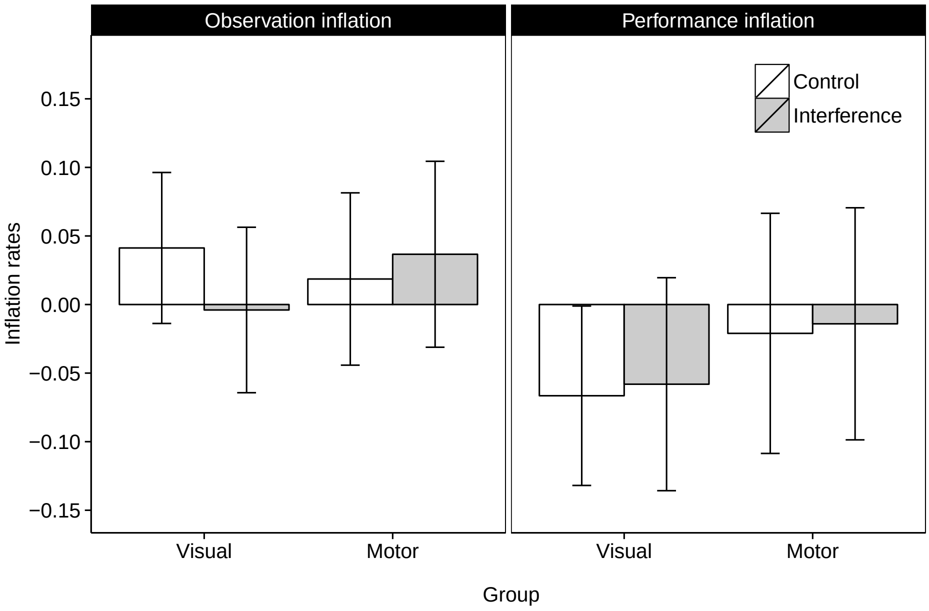

Figure 3.1. Observation and performance inflation by experimental group and interference condition in Experiment 4. Error bars are 95% between-subjects confidence intervals.

separately for the Visual and Motor group.

I first looked at observation inflation (false ‘Own’ responses to observed actions). For a significant inflation effect, I am looking for false ‘Own’ responses being given more frequently in response to observed than in response to novel actions, i.e., a main effect of Response type. In the Visual group, there were no significant main effects of Response Type or Interference condition, $Fs < 1$. There was no significant interaction, $F(1,22) = 1.59, MSe = 0.01, p = .22, \eta_{\mathrm{p}}^2 = .067$. In the Motor group there was no significant main effect of Response type, $F < 1$, but there was a significant effect of Interference condition , $F(1,22) = 5.52, MSe = 0.004, p = .028, \eta_{\mathrm{p}}^2 = .201$. False ‘Own’ responses occurred less frequently for actions retrieved under motor interference. There was no significant interaction, $F < 1$.

I next looked at false ‘Partner’ responses. Again, an inflation effect is characterised by a higher rate of ‘Partner’ responses to performed actions than to novel actions. In the Visual group, there was a significant main effect of Response type, $F(1,22) = 4.73, MSe = 0.02, p = .041, \eta_{\mathrm{p}}^2 = .177$. False ‘Partner’ responses were given more frequently to novel than to performed actions. There was no significant main effect of Interference condition, nor was there a significant interaction, $F < 1$. In the Motor group, there were no significant effects of Response type or Interference condition, nor was there a significant interaction, $Fs < 1$.

3.5.3 Discussion of model-free analysis

I analysed three aspects of the data with separate analyses: item memory, source discrimination and observation inflation. I will first discuss the item memory findings.

Item memory

The analysis of item memory shows three things. First, I observed a clear enactment effect. In line with the literature on enactment, performed actions were correctly recognized more frequently than observed actions (e.g. Engelkamp et al., 1994), regardless of interference condition or experimental group. This was still the case when participants' false alarm rates were taken into account (e.g. Golly-Haring & Engelkamp, 2003).

Second, occupying participants' motor system at recognition in fact improves their item recognition, in particular of performed actions. A caveat here has to be that examining the pattern of data between Visual and Motor group seems to suggest that this effect is due to a difference in responding to performed actions in the control condition rather than the interference condition. Given that I did not predict the observed effect of motor interference on item memory, I will refrain from attempting to interpret the benefit of motor interference on discrimination of old from new actions.

Third, participants were more likely to falsely classify novel actions as observed than performed (the 'It had to be you' bias). Interestingly, the rates of false alarms are considerably higher than the ones reported in a similar experiment by Hornstein and Mulligan (2004). Hornstein and Mulligan asked participants to perform and observe actions during encoding, and at retrieval complete a 3AFC (performed, observed, novel) source recognition task. Comparing participants' item memory performance in Experiment 4 (Table 3.2) with the same measures in Hornstein and Mulligan (Table 1) shows two things. First, participants' hit rates are comparable across the two experiments. Second, the false alarm rate in my experiment is 3 to 4 times the one that Hornstein and Mulligan report. Similarly, Lindner et al. (2010) report rates of false performed responses to novel items at below 4%. The comparable rates in my experiment are from 14% upwards.

There is a simple reason why participants are so much more likely to falsely accept novel items as old in the present experiment: the novel items I presented to participants are very similar to old items. Typically, the stimuli used in action memory experiments are a list of diverse objects coupled with unique actions (e.g. in Hornstein & Mulligan, 2004;

Leynes & Kakadia, 2013; Lindner et al., 2010). Objects are only used once throughout the experiment, and the actions paired with the objects are typical actions associated with the objects (though see Rosa & Gutchess, 2011). For any given participant, a random subset of actions is shown to them at encoding, with the remainder shown as distractors at test. No object appears twice, and consequently no actions appear twice (with some exceptions as the concept and motor action of some actions can be similar, i.e. 'Put an egg in the egg cup' is somewhat similar to 'Put a toothbrush in the cup'). To perform well at Old/New recognition, it is sufficient for participants to correctly recognize the objects. Items can be rejected as novel when the object itself is not familiar, without the motor component being taken into account.

In Experiment 4 the distractors were created using the same set of objects as the encoded items. The high false alarm rates in the experiment suggests that participants continued to make recognition decisions based on object-identity even though that was not diagnostic of the old-ness of an item. This is in line with results reported by Engelkamp and Zimmer (1995). They presented participants with simple verb-object phrases. They manipulated the distractors by creating similar and dissimilar distractor by exchanging the verb or the object of an encoded phrase. While similar distractors increased false alarm rates overall, this was particularly pronounced with a change in motor action when the object of the distractor remained the same as that of an encoded verb-object phrase.

The more sophisticated aspect of the creation of distractors, matching them to object combinations each participant had actually used in the experiment, did not selectively increase responding in line with the match. This suggests that re-using previously used objects at all was sufficient to disrupt Old/New recognition. Likely participants' memory for performed or observed actions was so poor to start with that they did not remember the object combinations they had used enough to be susceptible to this aspect of the manipulation. What this suggests is that participants' item memory performance in, for example, Hornstein and Mulligan (2004) may be strongly driven by object identification rather than motor action identification. Certainly, my experiment showed that discrimination of old and novel actions suffers greatly when participants' cannot discriminate those items simply on the basis of the objects.

A final point that bears mention that may have influenced false responding at test is that participants responded to the experimenter's verbal description of the actions. Since participants did not verbalize actions themselves, the experimenter verbally coded the actions (silently) during the encoding phase. It is possible that participants internally coded an action as ‘Push’ while the experimenter coded the same action as ‘Tap’ or ‘Shove’. Participants may have thus rejected actions they performed or observed as novel, since the specifics of the action kinematics did not match their memory. However, distractors were created using categorically different action verbs, i.e., the distractor to ‘push the mug with the card’ was ‘Put the card on the mug’ and not ‘Shove the mug with the card’. The process of verbal coding is therefore unlikely to have affected item recognition at test.

Source memory

I next looked at participants’ source memory. My key predictions for the effects of visual and motor interference at retrieval focused on participants’ source memory performance. Specifically, that motor interference should improve correct identification of observed and potentially disrupt identification of performed actions. Visual interference, on the other hand, should disrupt discrimination of observed actions.

There was no evidence for any effects of Interference in the source memory data, contrary to predictions. There are two possible explanations.

First, the low cognitive load of the interference tasks (watching flickering faces and moving the arms) may have been simply not strong enough to lead to interference. In Chapter 2, I manipulated the encoding of items, with the aim of selectively interfering with the encoding of specific memory traces. The encoding manipulation did in fact disrupt encoding but did not appear to *selectively* disrupt encoding. The approach in this experiment was instead to selectively interfere with the retrieval of those specific memory traces to ensure that participants had, in fact, encoded the items. There was no evidence for a disruption of source retrieval by interference. While Craik (2014) suggested that retrieval interference leads to smaller disruption than encoding interference on item memory measures, retrieval interference with a digit monitoring task has been shown to affect source monitoring performance (e.g. Dodson et al., 1998; Troyer & Craik, 2000). I chose the low-cognitive load to avoid generic disruption of source retrieval as shown by, for example, speeded responding (e.g., Benjamin & Craik, 2001). Yet, it is possible that not interference with the retrieval of source features, but only interference with the integration and processing of source information at retrieval can disrupt source monitoring

(Craik, Govoni, Naveh-Benjamin, & Anderson, 1996).

Second, participants' responding may not have been based on motor and visual information at all. As in the experiments in Chapter 3, participants generated the actions they performed. Self-performed actions could now be easily identified by the cognitive trace associated with them from the idea generation process. Observed actions were characterised by an absence of this cognitive trace. Correct source decisions about an action could now be made solely on the basis of the absence or presence of this cognitive trace of having generated these items. This means participants could ignore the motor or visual trace associated with the actions to come to correct decisions, and retrieval interference with those non-diagnostic traces would not impact responding. I will explore the impact of generation of actions versus mere production of actions on instruction in Chapter 4.

Inflation

Finally, I looked at the inflation effects in this experiment. I adapted the metric used by Lindner et al. (2010) to the present paradigm. Typically inflation effects occur when observing or imagining actions, potentially because enoding increases familiarity for or fluency of those actions relative to the actions that are not observed or imagined (Goff & Roediger, 1998). Increased false responses to observed or imagined actions now occur because participants misattribute the familiarity to the incorrect source at test (though see Thomas et al., 2003, for an alternative explanation).

In Experiment 4, I did not observe the observation inflation effect or the complementary performance inflation (increased rates of false 'observed' responses after performing actions). Rates of false responding to actions that had been encoded (old actions) in this experiment were comparable with Lindner et al. (2010), suggesting participants did commit source misattributions. On the other hand, as discussed above, false responses to novel items at test were higher than in Lindner et al. and cancelled out any inflation effects. In other words, false responding was already so high in this experiment that additional exposure to the item at encoding did not additionally lead to inflation of those responses. This does suggest that the familiarity of objects alone may drive the inflation effect in Lindner et al. (2010), with false source response rates only low when responding to objects that have not been shown previously.

3.5.4 Model-based analysis

I am centrally interested in participants' source memory performance. Throughout my thesis I am testing if source confusion of actions follows predictions of the source monitoring framework (Johnson et al., 1993) by looking at participants' source memory performance in the experiments. As I discussed earlier, while estimating source discrimination conditional on correct item recognition appears to give an unbiased estimate of source performance, this is not the case (e.g. Batchelder & Riefer, 1990; DeCarlo, 2003; Murnane & Bayen, 1996). The alternative to that model-free analysis is to estimate independent parameters for item memory, source memory and guessing or bias by fitting appropriate models to the data. While Murnane and Bayen (1996) suggested to fit multinomial processing models, in particular the Two-High-Threshold model (Bayen et al., 1996), an alternative is to fit signal detection models (DeCarlo, 2003).

While threshold models assume responding based on discrete states that result in rectangular memory distributions, responding in signal detection models is based on continuous memory distributions on a strength dimension. There is contention in the literature which model's assumptions better represent the data (Dube, in press; Pazzaglia et al., 2013; Slotnick & Dodson, 2005). Thus the majority of source memory modelling work has focussed on the curvature of ROC curves and slopes of zROCs for confidence rating data (for a review see Kellen & Klauer, 2015). Both models make different predictions given their assumptions about the distributions underlying performance (e.g. Batchelder & Alexander, 2013; Bröder & Schütz, 2009; Pazzaglia et al., 2013; Slotnick & Dodson, 2005; Yonelinas & Parks, 2007).

Since I did not collect confidence rating data, I will not comment on the underlying shape of the memory distributions in this experiment. Instead, I will fit both multinomial and signal detection models to the response selection data in the 3AFC source recognition task participants completed in this experiment.

I will fit three models in total: the Two-High-Threshold model (2HTM), the two-dimensional signal detection model (2D-SDT) and the one-dimensional signal detection model (1D-SDT). 2HTM and 2D-SDT map onto a source monitoring account of source memory performance (Johnson et al., 1993) by assuming separable item memory and source memory distributions. 1D-SDT maps onto a more simple relative memory strength account of source memory performance. In this account, responding is assumed to occur on the basis of overall memory strength alone (e.g. Marsh & Bower, 1993).

In the next section I will first determine the best-fitting model (i.e., the parameter restrictions that best account for the data) of each of these three types of model. To do this I will test models with different parameter restrictions against one another to compare them on their fit (by G^2, the negative log likelihood). Once I identified the winning model for each type of model, I will then compare those best-fitting models against one another to determine which type of model best accounts for the present data.

Model fitting

Two-High-Threshold model I first looked at the multinomial processing approach to source recognition data, pioneered by Batchelder and Riefer (1990) and Riefer, Hu, and Batchelder (1994), as an alternative to the estimation of source memory performance by rates of responding in source recognition experiments. The advantage of this approach is that source parameters are estimated independently of item memory and guessing parameters, so are free from biases and easily comparable across experiments.

The Two-High-Threshold model (2HTM) of source monitoring assumes separate decisions spaces for item memory and source memory components of the source monitoring task. Since both decisions separately are 2 alternative forced choice decisions (Old/New for item memory and Performed/Observed for source memory), both decision spaces are divided by two thresholds into three discrete areas (corresponding to latent states). In the item decision space, there is a detected-as-old area, a detected-as-new area and an undetected area. Only old items can cross into the detected-as-old area and only new items can cross into the detected-as-new area. The source decision space is divided up analogically into identified-as-performed, identified-as-observed and not-source-identified areas.

In the item decision space an item presented at recognition is designated as Old if it crosses the ‘Old’ threshold into the detected-as-old state and it is designated as New if it crosses the ‘New’ threshold into the detected-as-new state. If an item does not cross either threshold and remains undetected, participants guess with a certain probability that an old item is ‘Old’ and a new item is ‘New’.

Similarly in the source decision space, items presented at test either cross thresholds to be identified by source, or their source remains unidentified. In that case, participants guess with a certain probability that an item was performed and with the complementary probability that an item was observed.

Figure 3.2 illustrates the 2HTM for two sources (performed and observed actions) and novel items presented at test with item detection, source identification and guessing parameters. Participants' final responses can now arise from a combination of detecting and identifying the correct states of presented items, or guessing them. The parameters in the figure indicate the probabilities of item detection and item guessing, source identification and source guessing. D_P and D_O indicate the probability with which performed and observed actions respectively are detected as old and D_N gives the probability of novel actions identified as new. If items are not detected successfully as old or new, they are now guessed to be old with probability b or new with complementary probability $1-b$. Performed or observed actions that were detected correctly as old, are now identified as originating from the correct source with probability d (d_P for performed and d_O for observed actions). Actions that were guessed to be old (rather than detected as old) are guessed to be performed with probability g and guessed to be observed with probability $1-g$. Performed and observed actions that were correctly identified as old but whose source was not detected are guessed to be performed with probability a and guessed to have been observed with probability $1-a$.

The probability of responses (on the right-hand side of Figure 3.2) can be calculated by multiplying probabilities along the branches of the tree and adding resulting probabilities across the branches of the tree. The probability of an 'Own' response to an action participants in fact performed themselves, for example, results from $D_P * d_p + D_P * (1-d_p) * a + (1-D_P) * b * g$, with three different branches in the first decision tree resulting in 'Own' responses.

Using the approach shown by Bayen et al. (1996), I estimated the best fitting model for participants' aggregate data, separately for the Visual and the Motor group (also see Bayen, Nakamura, Dupuis, & Yang, 2000; Schütz & Bröder, 2011). I describe the detailed model-fitting procedure in Appendix B.2.1.

At the top of this chapter I predicted that participants would have higher item memory for performed than for observed actions (the enactment effect). In the 2HTM this should be reflected in a higher item parameter for performed than observed actions. I also predicted that interference would affect source memory for observed actions in particular, or bias responding towards them. In the 2HTM this should be reflected in an effect of interference

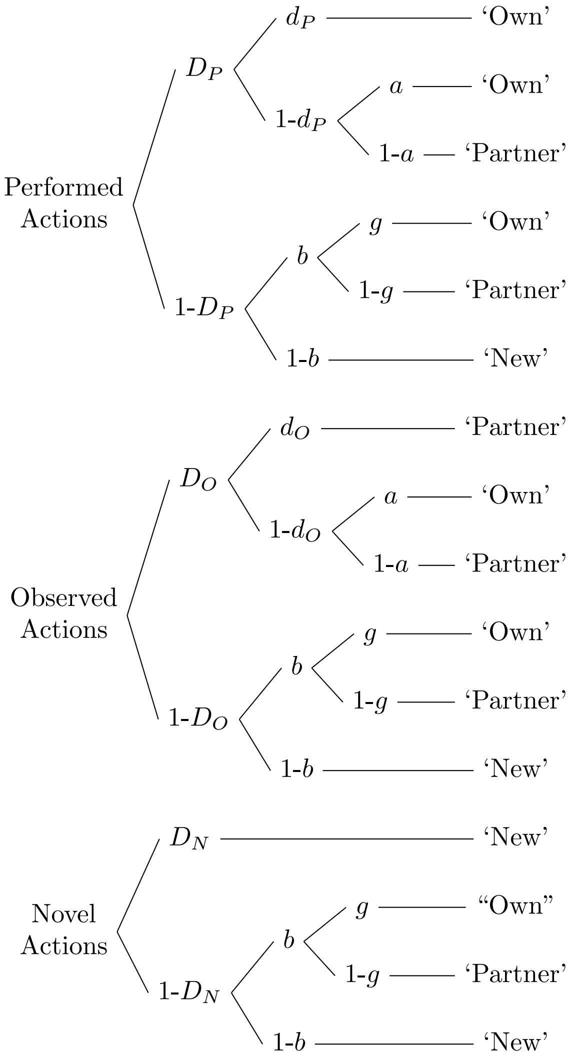

Figure 3.2. Two-high-threshold multinomial model of source monitoring for two sources with the first source performed, the second observed actions. D_P = probability of detecting a performed action; D_O = probability of detecting an observed action; D_N = probability of detecting the item is new; d_P = probability of correctly discriminating the source of a performed action; d_O = probability of correctly discriminating the source of an observed action; a = probability of guessing that a detected item was a performed action; b = probability of guessing an item is old; g = probability of guessing that an undetected item was a performed action

Table 3.6. Parameter estimates for the best-fitting Two-High-Threshold model in Experiment 4

	D_P	$D_O = D_N$	$d_P = d_O$	b	$a = g$
Visual group					
Control	.61[.55,.68]	.29[.24,.33]	.83[.74,.93]	.60[.57,.63]	.34[.30,.37]
Interference	.61[.55,.68]	.29[.24,.33]	.83[.74,.93]	.60[.57,.63]	.34[.30,.37]
Motor group					
Control	.54[.47,.61]	.27[.23,.32]	.77[.66,.87]	.58[.55,.61]	.41[.38,.45]
Interference	.54[.47,.61]	.27[.23,.32]	.77[.66,.87]	.58[.55,.61]	.41[.38,.45]

Parameter estimates with 95% confidence interval in brackets.
D_P, D_O, D_N - item memory parameters
d_P, d_O - source memory parameters
b - item (Old/New) guessing
a, g - source (Performed/Observed) guessing

on source memory and/or guessing parameters.

Table 3.6 shows the parameter estimates of the best-fitting 2HTM model, identified by likelihood-ratio tests. Parameters can range from 0 to 1. Item recognition parameter (D) estimates higher than 0 indicate that some item recognition is present. Source memory parameter (d) estimates higher than 0 indicate that some source memory is present. Source guessing parameters (a and g) estimates higher than the chance level of 0.5 indicate a guessing bias towards 'Own' responses, while estimates lower than 0.5 indicate a guessing bias towards 'Partner' responses. Finally item guessing parameter (b) estimates higher than 0.5 indicate a guessing bias to judge an item as old while estimates lower than 0.5 indicate a guessing bias to judge an item as new.

The winning model is characterized by 5 parameters. The parameter estimates show clear evidence for an enactment effect, with the probability of detecting a performed action (D_P) as old higher than the probability of detecting an observed action as old (D_O) in both the Visual and the Motor group. Across source parameters and guessing parameters, there was no evidence for an effect of interference. Participants were biased to guess items as 'Old' overall (with b exceeding 0.5). While this may be due to the high similarity of distractor items, item guessing parameters of similar size have been reported in the literature (Bayen et al., 2000; Schütz & Bröder, 2011). Participants overall were also biased to respond 'Partner' (with a and g not exceeding 0.5), in part reflecting the 'It had to be you' bias.

Signal Detection Models

The basis of signal detection theory is that types of items can be represented by any, typically normal, distributions in one or multi-dimensional space. Participants set re-

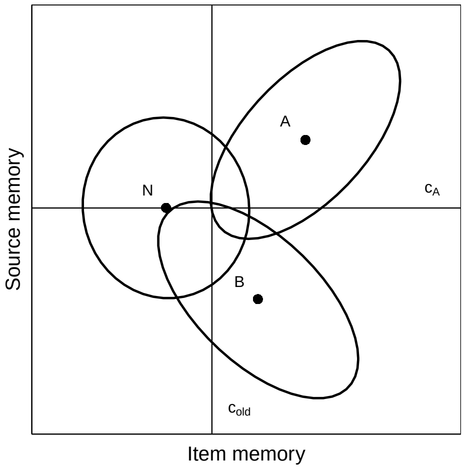

Figure 3.3. Illustration of the two-dimensional signal detection model for separate item and source memory dimensions for novel items (N) and two sources of old items (A and B), with decision criteria each on the item memory dimension (c_{old}) and the source memory dimension (c_A)

sponse criteria on each dimension. Signal detection modelling is typically used to model data from uncertainty estimation (i.e. confidence ratings) rather than response selection paradigms. The response selection case is a reduced implementation of the multiple-criteria idea. DeCarlo (2003) formulated the model for this case.

Modelling source monitoring data with signal detection approaches provides information on participants' ability to discriminate sources (how far distributions are apart), if item and source memory performance are correlated (correlation between dimensions) are and whether and how participants are biased in their responding on the various dimensions.

Two-dimensional signal detection model Following predictions of the source monitoring framework (Johnson et al., 1993), source recognition is typically modelled using two-dimensional signal detection models. Here, separable item memory and source memory dimensions are assumed, with participants setting criteria on both dimensions and making appropriate responses if an item's strength exceeds or does not exceed the criteria. An illustration of the two-dimensional model for item and source memory dimension is given in Figure 3.3.

When a response is given under this model it is classified as old (i.e., A or B) if its value of item strength exceeds the c_{old} threshold, else it is classified as 'New'. If the value of the item's source strength exceeds the c_A value, it is classified as 'A', else as 'B'. In the 3AFC case, decisions on the source dimension are contingent on an 'old' decision on the item dimension. In my case A represents performed actions and B observed actions. When the c_{old} threshold is exceeded participants respond 'Own' or 'Partner' in response

Table 3.7. Parameter estimations the best-fitting 2D-SDT model in Experiment 4

	D_P	D_O	d_P	d_O	c_{old}	c_{own}
Visual group						
Control	1.00 [.97,1.03]	.67 [.55,.78]	1.00 [.97,1.03]	0.00 [-.00,.00]	.10[.03,.16]	.51[.44,.57]
Interference	1.00 [.97,1.03]	.67 [.55,.78]	1.00 [.97,1.03]	0.00 [-.00,.00]	.10[.03,.16]	.51[.44,.57]
Motor group						
Control	1.00 [.95,1.05]	.68 [.57,.79]	.99 [.85,1.13]	0.00 [-.00,.00]	.17[.11,.24]	.42[.33,.51]
Interference	1.00 [.95,1.05]	.68 [.57,.79]	.99 [.85,1.13]	0.00 [-.00,.00]	.17[.11,.24]	.42[.33,.51]

Parameter estimates with 95% confidence intervals in brackets
D_P, D_O - memory strength of performed and observed actions relative to novel items on the item memory dimension
d_P, d_O - memory strength of performed and observed actions relative to novel items on the source memory dimension
c_{old} - decision criterion on the item memory dimension
c_{own} - decision criterion on the source memory dimension

to items, when additionally the c_{own} threshold is exceeded, participants classify items as 'Own', otherwise as 'Partner' items.

For the purposes on this analysis, I will use the same terms to describe item memory and source memory strength as in the 2HTM model fitting section above. Typically, the memory strength of novel items is set to 0 on both memory dimensions and the memory strength of performed and observed actions is estimated relative to 0. I will call the average item memory strength of performed actions D_P and of observed actions D_O. The average source memory strength of performed actions will be designated d_P and of observed actions d_O.[2]

I fitted the equal-variance (all distributions are set to have a variance of 1) two-dimensional signal detection model with the *MPTinR* package (Singmann & Kellen, 2013). The model was fit separately to the Visual group and Motor group. I describe the detailed model-fitting procedure in Appendix B.2.2. I assumed equal-variance and independence of dimensions to limit the number of parameters that were fitted. The average memory strength of novel items was assumed to be 0 on both the item memory and source memory dimension, with the distribution assigned a standard deviation of 1 for both dimensions. Given the theoretical predictions, I tested whether motor and visual interference would results in shifts in response criteria on the item and source memory dimension.

Table 3.7 shows the parameter estimates for the best-fitting 2D-SDT (equal-variance, independent dimensions) model.

The parameter estimates of the winning 2D-SDT model show clear evidence of the enactment effect, with performed actions showing higher average memory strength than

---

[2] Note that the parameters here indicate average memory strength not probabilities.

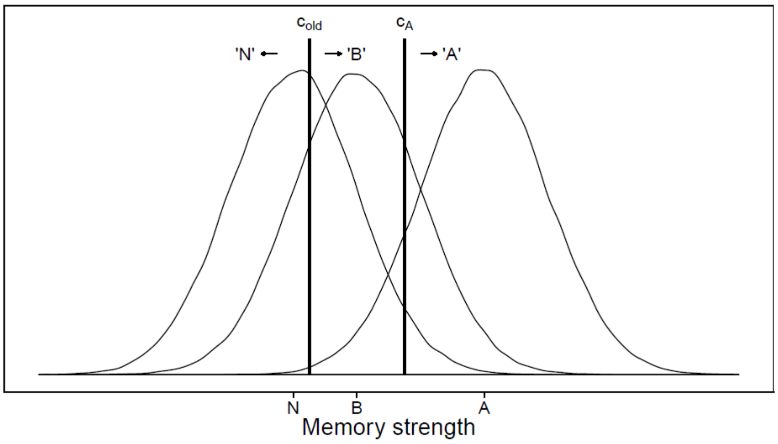

Figure 3.4. Illustration of the one-dimensional signal detection model of source memory for novel items (N) and two sources of old items (A and B), with decision criteria are c_{old} andc_A

observed actions. The relative order of the item memory strength distributions (novel, observed, performed) can account for the ‘it had to be you’ bias under Hoffman (1997)’s relative strength account. Participants are able to distinguish performed from observed actions, though performance in the task can be accounted for by not assuming any source memory strength for observed actions. There was no evidence for interference affecting performance in the task such as the c_{own} criterion shifting to allow more ‘Partner’ responses.

One-dimensional signal detection model The two-dimensional signal detection model of source memory assumes that distributions of items in a source memory experiment are best described by assuming separate item memory strength and source memory strength dimensions. Decisions about the old-ness of an item depend on a criterion set on the memory strength dimension, while further decision about the source-belongingness of an item depends on a criterion set on the source memory strength dimension.

Given the discussion about dual- versus single-process models in learning, memory and decision making (e.g. Berry, Shanks, Speekenbrink, & Henson, 2012; Nosofsky & Kruschke, 2001; Shanks & Berry, 2012), the obvious alternative to the complex two-dimensional signal detection model is a one-dimensional signal detection model. The theoretical assumption of such a model is that the source dimension cannot be separated from overall memory strength. Under this account, judgements about items are made on the basis of the overall strength of an item, and performance in a source monitoring task arises from the difference in the relative strengths of the distributions. This model is illustrated in Figure 3.4.

Table 3.8. Parameter estimations for best-fitting 1D-SDT model in Experiment 4

	D_P	D_O	σ_P	σ_O	c_{old}	c_{own}
Visual group						
Control	1.87 [1.64,2.10]	.51 [.40,.62]	1.76 [1.50,2.02]	1.00 *	.07 [-.00,.16]	1.31 [1.21,1.41]
Interference	1.87 [1.64,2.10]	.51 [.40,.62]	1.76 [1.50,2.02]	1.00 *	.07 [-.00,.16]	1.31 [1.21,1.41]
Motor group						
Control	1.45 [1.27,1.64]	.48 [.37,.59]	1.57 [1.34,1.80]	1.00 *	.10 [.02,.18]	1.15 [1.06,1.25]
Interference	1.45 [1.27,1.64]	.48 [.37,.59]	1.57 [1.34,1.80]	1.00 *	.10 [.02,.18]	1.15 [1.06,1.25]

Parameter estimates with 95% confidence intervals in brackets
D_P, D_O - memory strength of performed and observed actions relative to novel items on the item memory dimension
$sigma_P, sigma_O$ - standard deviations of performed and observed actions distributions
c_{old} - Old/New decision criterion
c_{own} - Performed/Observed decision criterion
* - fixed parameter

In the present experiment Source A are performed actions and Source B are observed actions. Participants will judge an item as 'Own', if its strength exceeds the decision criterion participants set for 'Own' responses (here: c_{own}). If an item exceeds the criterion for old (c_{old}) responses but does not exceed c_{own}, participants will judge the item as 'Partner'. I will re-use the terminology for the previous model fitting procedures to call the average strength of performed actions D_P and the average strength of observed actions D_O, with both estimated relative to novel actions at 0.

I fitted equal-variance and unequal-variance one-dimensional signal detection models with the *MPTinR* package (Singmann & Kellen, 2013). The models were fit separately to the Visual group and Motor group. I describe the detailed model-fitting procedure in Appendix B.2.3. The mean of the distribution of novel items was set at 0, with a standard deviation of 1. Given the theoretical predictions, I tested whether motor and visual interference would results in shifts in response criteria on the item and source memory dimension.

Table 3.8 shows the parameter estimates for the best-fitting 1D-SDT model for the Visual and Motor group. I predicted that interference conditions would affect the placement of decision criteria in this model. The parameter estimates show clear evidence of an enactment effect, with performed actions characterized by higher memory strength than observed actions. The memory strength of performed actions is more variable than that of observed actions. This finding mirrors results of higher variability for target than lure distributions (Mickes, Wixted, & Wais, 2007). This may reflect that the enactment effect (i.e. the resulting higher quantity of retrieved performed actions than observed actions) may be driven in part by the vastly higher memory strength of some performed actions in addition to an on average higher memory strength of performd action. As with the previous models, there was no evidence for an effect of interference, and the best fitting model estimated parameters across the control and interference conditions in both the Visual and Motor group.

Comparison of model fits

The experimental manipulation I tested in Experiment 4 was retrieval interference (visual or motor retrieval interference). The model-based analysis did not diverge from the model-free analysis. Across all models, there was no evidence for an effect of either visual or motor interference. All models provided best fits with parameters were estimated across the control and interference conditions. I will now move on to compare which of these models accounted best for the data.

In the previous section I fitted 2HTM, 2D-SDT and 1D-SDT models to the data. When comparing the models for overall best model, I am particularly interested in two comparisons (1) comparing 2HTM and 2D-SDT and (2) comparing 2D-SDT and 1D-SDT.

For (1), both 2HTM and 2D-SDT represent performance in a source monitoring task as based on judgements of items on different item memory and source memory dimensions. Both assume participants' source judgements are more or less independent of participants' item recognition for each item, in line with a source monitoring account (Johnson et al., 1993). The critical difference between the models are the assumptions about the shape of the memorial distributions underlying performance. While the 2HTM assumes discrete states resulting in rectangular distributions, the 2D-SDT model assumes that both on the item memory and source memory dimension, memory strength is normally distributed from less to more memory strength. I did not test the distributional assumptions of both models in this experiment as typically done when testing which model is better suited to account for the data (e.g. Batchelder & Alexander, 2013; Bröder & Schütz, 2009; Pazzaglia et al., 2013; Slotnick & Dodson, 2005; Yonelinas & Parks, 2007). Nevertheless, I can still compare which type of model (each type represented by its best-fitting instantiation) does better in accounting for the 3AFC source recognition data, given its fit and number of parameters (see Schütz and Bröder (2011) for a complementary approach in fitting 2HTM to confidence rating data). This comparison therefore is technical, rather than theoretical. Given the same theoretical assumptions, which of the two mathematical instantiations of this theoretical idea is better suited to account for this particular set of data.

For (2), both the 2D-SDT and 1D-SDT model assume the same type of underlying distributions. They critically differ in the complexity of the model. While the 2D-SDT model assumes separate dimensions for item and source memory, in line with the source monitoring account (Johnson et al., 1993), the 1D-SDT model assumes that performance in a source monitoring task can be accounted for by distributions on a single overall memory strength dimension. Under this account, participants do not evaluate source features at test, but base their source monitoring responding on judgements of the relative strength of items (Hoffman, 1997; Marsh & Bower, 1993). This comparison is theoretical, rather than technical. Given different theoretical assumptions, which model better accounts for the data.

Figure 3.5 shows the data observed in the experiment and predicted by the models for the Visual group and Figure 3.6 shows the same for the Motor group. The data is plotted as the proportion of responses (‘Own’, ‘Partner’, ‘New’) given to each source of items (Performed, Observed, Novel). Visual inspection of the predictions evidences how comparatively well all models accommodate the data, with the majority of predicted data points falling within the 95% confidence interval. Visual inspection suggests that 2HTM and 2D-SDT model provide comparative, good fits to the data. The 1D-SDT model struggles to accommodate some aspects of the data, in particular responses to observed actions, though note that the predictions are based on the parameters for the aggregate data.

Table 3.9 shows the formal comparison of all three models. I compared the best-fitting 2HTM, 2D-SDT model and 1D-SDT model using the *MPTinR* package (Singmann & Kellen, 2013). The table lists AIC and BIC alongside the model fit G^2 (for the negative log likelihood). BIC is defined as $BIC = k * ln(n) - 2 * ln(L)$ where k denotes the number of parameters, n the number of observations and L the maximized value of the likelihood function. The BIC is used when the aim of the model fitting is to maximize fit while minimizing parameters, with a lower BIC indicating a better model fit giving parameters and number of observations. I also calculated AIC values, they are listed in Table B.3 and lead to the same conclusions.

Table 3.9 shows for (1) that the 2HTM clearly outperforms the 2D-SDT model in accommodating the data. For the aggregate data, 2HTM results in a lower BIC than the 2D-SDT model. Even when comparing them on how well they account for each individuals’ data,

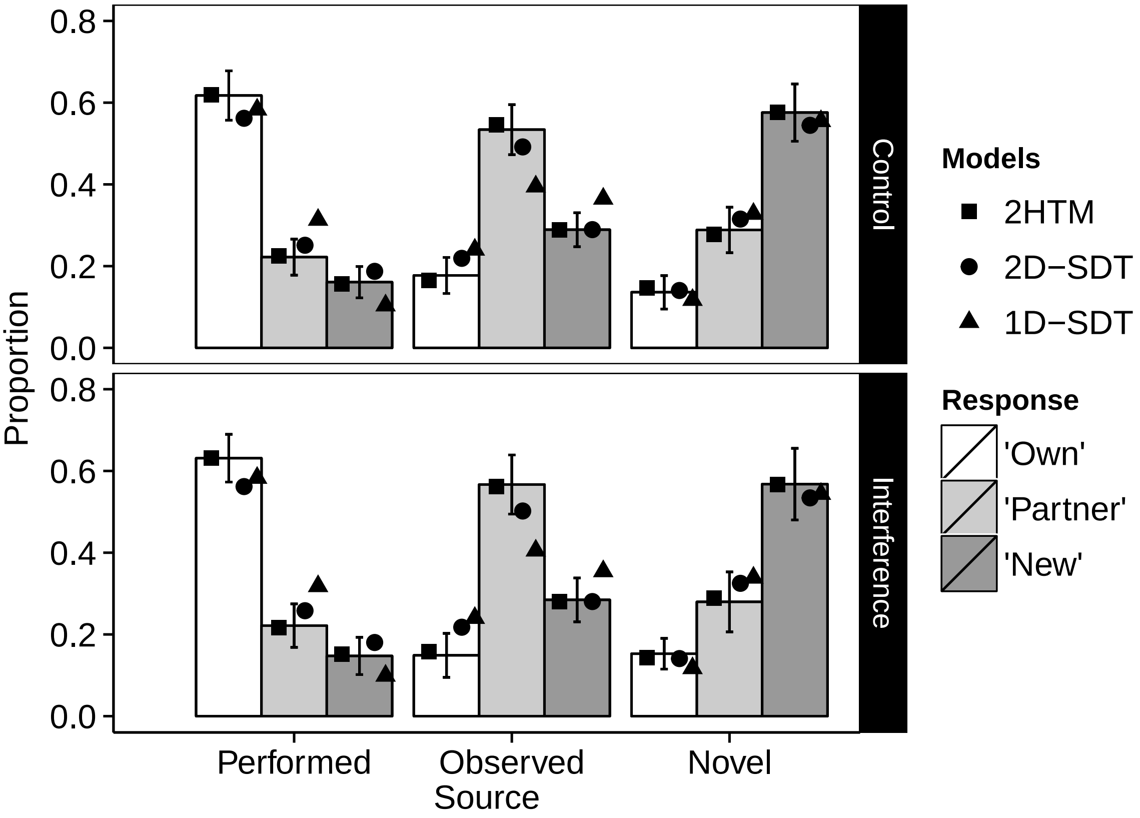

Figure 3.5. Relative proportions of responses to item types in Experiment 4 in the Motor group, with the observed data shown in the bars. The error bars are 95% confidence intervals. The model predictions are based on the aggregate data. The Two-High-Threshold model is denoted by the squares, the two-dimensional signal detection model by the filled circles and the one-dimensional signal detection model by the open circles

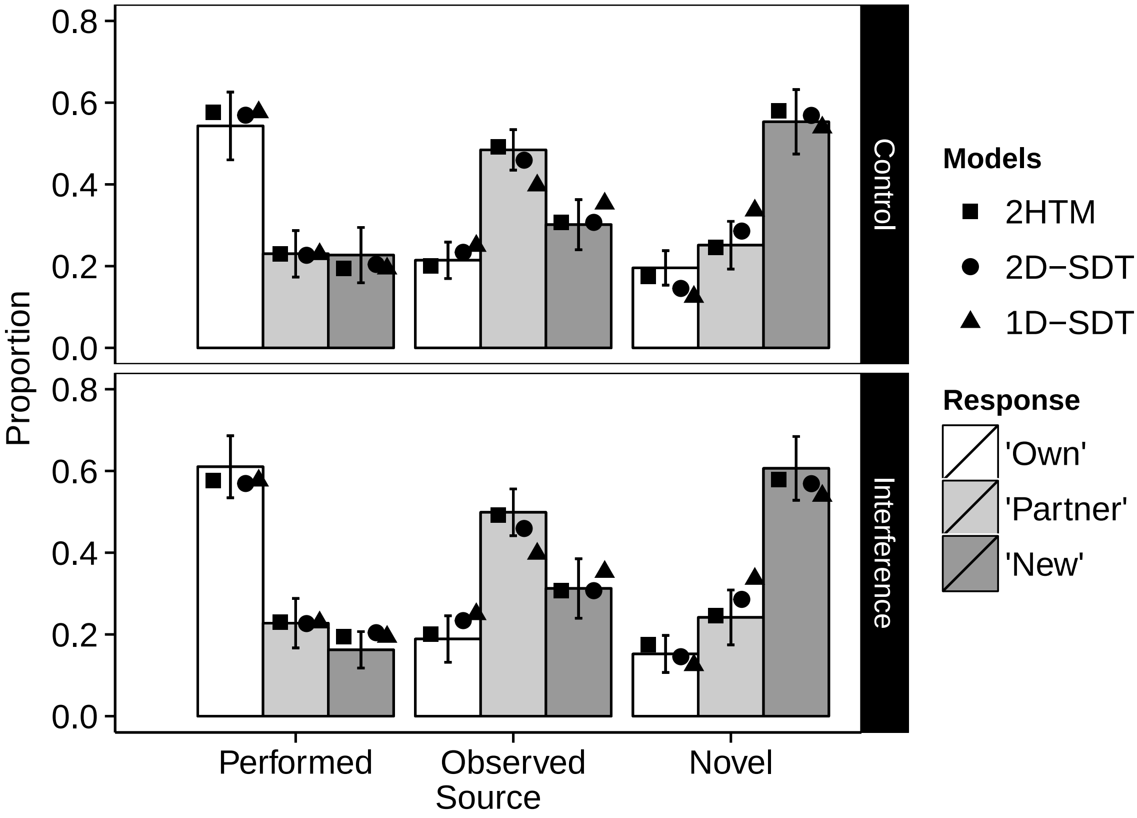

Figure 3.6. Relative proportions of responses to item types in Experiment 4 in the Motor group, with the observed data shown in the bars. The error bars are 95% confidence intervals. The model predictions are based on the aggregate data. The Two-High-Threshold model is denoted by the squares, the two-dimensional signal detection model by the filled circles and the one-dimensional signal detection model by the open circles

Table 3.9. Comparisons of model fits of the 2HTM, 2D-SDT and 1D-SDT models in Experiment 4

		Individual data					Aggregate data		
Model	Parameters	G^2	p	best describes		BIC	G^2	p	BIC
Visual group									
2HTM	5	213.47	.003	23		1110.72	2.21	.95	41.22
2D-SDT	6	364.94	<.001	0	2	1441.64	39.10	<.001	85.91
1D-SDT	5	295.50	<.001		21	1192.76	74.86	<.001	113.87
Motor group									
2HTM	5	190.11	.06	23		1088.77	10.18	.18	49.25
2D-SDT	6	308.19	<.001	0	2	1286.58	25.60	<.001	72.49
1D-SDT	5	260.54	<.001		21	1159.20	79.15	<.001	118.22

G^2 - goodness of fit BIC - Bayesian Information Criterion

the 2HTM better accounts for the data than 2D-SDT model for 23 out of 23 participants.

Comparing the two signal detection models against one another for (2) leads to less clear results. Both models differ significantly from the observed data (indicated by significant G^2 values). For the aggregate data, 2D-SDT is preferred by BIC. However, 1D-SDT, despite its lower complexity and fewer parameters, is better able than 2D-SDT to account for a majority of participants in both the Visual and Motor group when looking at the fits to individuals' data.

3.5.5 Discussion model-based analysis

I conducted a model-based analysis for two reasons. First, the model-based analysis allowed isolated estimation of item and source memory parameters, avoid the conditional estimation of source memory performance on item memory performance (e.g. Murnane & Bayen, 1996). This analysis complemented the results of the model-free analysis. There was no evidence for the predicted effects of visual or motor interference on performance. Throughout the model fitting procedures for all models, there were some hints about participants' item memory for performed actions improving under motor interference, but the effect was not clear enough to warrant inclusion in the best-fitting model for any of the three types of models.

The second aim of the model-based analysis was to (a) compare performance of threshold and signal detection models for response selection data in a 3AFC source recognition test of action memories, and (b) compare performance of two-dimensional and one-dimensional signal detection models on fitting source recognition data since they make different theoretical claims about the basis of participants' source judgements.

For (a), the Two-High-Threshold model clearly out-performed the signal detection models in fitting the aggregate and individuals' data. While the Two-High-Threshold model and two-dimensional signal detection model are both based on the assumption that memory for an item and memory for its source are based on two separate dimensions, they make diverging assumptions about the shape of the underlying distribution. While it is tempting to argue that the assumption of rectangular distributions and threshold processes then clearly provides a better account of the data, that may be premature.

For one, I did not explicitly test what the shape of participants' underlying memorial distributions is. Typically this is done by collecting confidence ratings and constructing a ROC curve (plotting hits versus false alarms for increasing confidence), with the signal detection model predicting a curvilinear form based on the continuous distribution assumptions and the threshold model predicting a straight line based on the rectangular distribution assumptions (e.g. Batchelder & Alexander, 2013; Bröder & Schütz, 2009; Pazzaglia et al., 2013; Slotnick & Dodson, 2005; Yonelinas & Parks, 2007). That analysis is beyond the scope of this project.

Second, the two-dimensional signal detection model I tested was severely restricted to allow fitting the limited number of observed data points. I assumed independent memory dimensions and equal-variance, for example. It is possible that the fit of, in particular individual participants' responses, would have improved with fewer parameter restrictions. Note, the parameter estimations (with means of the memory strength distributions estimated around 1 and 0 with little flexibility) seem to suggest that the fitting procedure itself (possibly given the already high number of parameters) may have not resulted in the best representation of the two-dimensional signal detection model. Though, the overall fit shown in the figures seem to suggest that those estimates can predict the data well.

My results are roughly in line with Schütz and Bröder (2011) and Bröder and Schütz (2009). They fit Two-High-Threshold models and signal detection models to confidence rating data (by allowing guessing parameters to vary across the confidence rating bins) and equally found that both models accounted for the data, with the Two-High-Threshold model edging out the signal detection model in some tests.

For (b), the one-dimensional signal detection model outperformed the two-dimensional signal detection model on the individual fits but not on the aggregate data. Wickens (2002) pointed out that in particular in the signal detection case, fitting aggregate data may skew parameter estimations and model fit given the non-linear relationship between hits and false alarm rates (also see Morey, Pratte, & Rouder, 2008). Lewandowsky and Farrell (2011) recommend, based on results by A. L. Cohen, Sanborn, and Shiffrin (2008), to base evaluation of the model fits on the individuals' model fit rather than the aggregate data's fit if the data do not point to the same conclusion. In that case, the modelling in this chapter provides some support that the data in a 3AFC source recognition experiment can be accounted for by a relative memory strength model of source memory (Hoffman, 1997; Marsh & Bower, 1993), at least when the source monitoring account is equally represented by a signal detection model. Contrary to the theoretical predictions of the source monitoring model, the mathematical model that assumes that participants judge source on overall memory strength rather than evaluation of features, provides a better account of participants' responding at test.

Again, it is possible that a less restricted two-dimensional signal detection model (especially given the gain in fit the unequal variance assumption for the one-dimensional model provided) may have provided a better fit. However, given the much higher parameter space of such a model, it is then questionable if that is economical (if one assumes that parameter minimization best represents cognitive processes). This finding is in line with recent advances towards a single-system account of memory (e.g. Berry et al., 2012; Shanks & Berry, 2012).

3.6 Discussion

Experiment 4 differed from Experiments 1 through 3 in experimental design, stimuli and retrieval task. As in the recall experiments in the previous chapters, I found a robust enactment effect (e.g. Engelkamp et al., 1994) in this source recognition paradigm. Participants were better able to identify performed actions as old than observed actions.

As in the recall experiments, I found evidence for the 'It had to be you' bias, where participants are more likely to determine novel items to belong to an external than an internal source (Bink et al., 1999; Hoffman, 1997; Johnson et al., 1993). This bias was evident in both participants' responding at test and reflecting in their source guessing (when fitting a Two-High-Threshold model) being biased towards 'Partner' responses.

In Experiments 1 through 3, I showed that participants misremember observed actions as self-performed (as in the observation inflation effect; Lindner et al., 2010), but that they also misremember self-performed actions as observed. Experiment 4 similarly showed that participants misattribute observed actions to themselves but also misattribute performed actions to have been observed. In line with other source recognition of action memory studies (Hornstein & Mulligan, 2004; Leynes & Kakadia, 2013; Rosa & Gutchess, 2011) this suggests that, contrary to motor simulation account predictions, source misattributions of action memories occur in both directions.

Critically, I tested whether selectively interfering with the retrieval of motor and visual memory traces would affect source memory performance. The prediction was based on the assumption that participants' source decisions are based on evaluation of qualitative memory features at retrieval (Johnson et al., 1993). Interfering with the reactivation of motor interference should have increased participants' source memory performance relative to baseline. Experiment 3 showed an effect of motor (though also high cognitive) interference at encoding limiting participants' subsequent source memory performance and retrieval of actions encoded under interference, but showed no evidence of a predicted beneficial effect of selective motor interference on source memory. Similarly, I did not observe a predicted beneficial effect of selective motor interference at retrieval on source monitoring performance. That means that this experiment, like the prior experiments, failed to provide evidence in line with a motor simulation account of source memory for actions.

In fact, looking at the pattern of responding in Experiment 4 and other experiments in the literature (Hornstein & Mulligan, 2004; Leynes & Kakadia, 2013; Lindner et al., 2010; Rosa & Gutchess, 2011) calls into questions if reactivation of the motor trace crucially guides responding at test. Typically stimuli in source recognition of action memory studies are lists of verb-object phrases (Hornstein & Mulligan, 2004; Leynes & Kakadia, 2013; Lindner et al., 2010; Rosa & Gutchess, 2011). Each phrase combines a new object with an action, so objects are not re-used in the experiment. Some of the list is shown at encoding, with the remainder of the list used as distractors. In Experiment 4, I used sets of objects at encoding, with distractors created by re-using objects from encoding. I showed increased false responding to distractors in Experiment 4 compared to the rates typically reported in the literature. This suggests that participants' source decisions about action memories are based, at least in part, on their familiarity with the object, without consideration of the motor trace, rendering an interference with the retrieval of that motor trace ineffective.

While this experiment did not provide evidence for a motor simulation account of source memory for actions (Lindner et al., 2010), it also did not provide any evidence for a source monitoring account of source memory (Johnson et al., 1993). Under the latter account the evaluation of source features separate from item memory guides participants' source responding. Under that account visual and motor interference should have interfered with participants' ability to use those source cues and make correct source decisions. There was no evidence for that. In fact, the formal model that predicts source responding on the basis of overall memory strength could account well for the data (Hoffman, 1997; Marsh & Bower, 1993). It is possible that the evaluation of source features at test is not critical for participants' source performance.

A second possibility is that the cognitive trace encoded by generating actions at encoding guided source responding at test (rather than the motor or visual trace). Participants generated all actions they performed and none of the actions they observed. Performed could be distinguished from observed or novel actions simply on the basis of participants deciding whether they generated the verb-objects phrase themselves or whether they had not, without taking performance of the action into account. I will explore the role of the cognitive trace in Chapter 4.

Chapter 4

Generation

4.1 Background

In Chapter 2 and 3, I tested whether motor representations encoded during observation could impair source memory performance by increasing participants' tendency to claim observed actions as self-performed. While those chapters focused on the motor trace as a feature potentially encoded during action observation, this chapter will focus on the cognitive trace, i.e. the cognitive operations encoded alongside the item memory such as generating the idea for an action. According to the source monitoring framework (Johnson et al., 1993), qualitative features are encoded alongside the memory for an item and the source of an item can be inferred from evaluation of those features at retrieval. Source errors occur if participants did not encode qualitative features, do not evaluate them at retrieval, or if features overlap to such a degree that they are not diagnostic of one or the other source. Mitchell and Johnson (2000) argued that participants determine the source of an event by comparing the qualitative features of that event to a set of diagnostic features that are prototypical for that kind of event (see also Leynes & Kakadia, 2013). Cognitive features are more diagnostic of internally-generated events than external events, while external events are associated with richer perceptual features. Generating actions should therefore be taken as a cue for an action having been performed, rather than observed.

In this chapter, I manipulate the strength of the cognitive trace associated with performed actions in Experiment 5 and Experiment 6, by asking participants to generate the ideas for the actions they perform. In Experiment 7, I manipulate the cognitive trace to be associated with both performed and observed actions, by asking participants to generate the idea for both actions they perform and actions they observe.

In experiments of source memory for actions, participants are typically shown a list of action phrases (e.g. Hornstein & Mulligan, 2004; Leynes & Kakadia, 2013; Lindner et al., 2010; Rosa & Gutchess, 2011). Participants perform some of these action phrases and observe others being performed. Participants' subsequent source decisions about an action having been performed rely on participants' memory for the perceptual features associated with both events, such as visual features, somatosensory or motor features. Cognitive features can at best play a minor role in participants' source decisions since they never generated the idea for an action. On the other hand, the experiments I have presented thus far contrasted participants' memory for actions they generated and subsequently performed with those they observed a partner perform. In those experiments, performed actions could be identified as performed in two ways: participants retrieved the motor features associated with performance or they retrieved the cognitive operations associated with generating the idea for an action.

This presents two possible scenarios for participants' performance. First, it is possible that participants made use of the additional information and considered both perceptuo-motor and cognitive features at retrieval. This would have made distinguishing performed and observed actions easier and improved source performance, since now performed actions are not only uniquely identified by a motor trace but also uniquely identified by having been generated by participants. Hornstein and Mulligan (2004) used non-generated action phrases in a recognition paradigm. The rates of false responses they report are comparable to those I reported, suggesting there is no clear absolute benefit of having generated the idea for an action on source memory. However, rates of false responding would be influenced not only by items having been generated or not, but also by, for example, list length, delay and similarity of items (Johnson et al., 1993), so equivalent rates here are not conclusive evidence against an additional benefit of generation of memory performance.

I will test whether participants are making use of the cognitive trace when it is available in Experiment 5 and Experiment 6. Here participants will be asked to make source decisions between actions they performed after generating the idea and observed actions (as in previous experiments) and make source decisions between actions they performed on instruction and observed actions.

A second possibility is that participants are not only making use of the cognitive trace, but that responding is based exclusively on it, while perceptuo-motor information is ignored. In the experiments I reported thus far, performed action were always self-generated. Participants' source responding could have occurred on the basis of this cognitive involvement with items alone. It is therefore possible that participants never considered the motor and perceptual features of an action at all and solely focused on the cognitive processes to distinguish performed or observed actions. I will test this in Experiment 7. Here participants generate actions that they then perform or observe their partner perform. Generation is now not diagnostic of action performance, forcing participants to take the remaining features into account.

4.2 Aims and predictions

The experiments in this chapter will investigate the impact generating the ideas for actions has on subsequent source memory for performed versus observed actions. Do participants systematically use the cognitive trace as a basis of their source decisions about performed and observed actions, and do they know to ignore that feature when it is no longer diagnostic of the source in the experiment or do they continue to use it because it is typically diagnostic? A secondary aim of this chapter is to disentangle different source memory accounts, in particular diagnostic features (source monitoring), mistaken fluency and relative strength accounts.

Diagnostic features account: This account assumes a dissociation of overall memory strength and source memory strength (akin to two-dimensional signal detection models) and predicts that participants source performance is based on systematic evaluation of source features and is separate of (though possibly correlated with) items' overall memory strength (Johnson et al., 1993). Under a diagnostic feature account, participants should be able to make use of the cognitive trace in addition to the motor trace created during performance of self-generated actions (Leynes & Kakadia, 2013; Mitchell & Johnson, 2000). They should then be more likely to claim those actions as self-performed relative to actions they only performed on instruction. This should increase participants' correct recall of performed actions and decrease the false recall of performed actions (leading to identical predictions as made by a heuristic account of source evaluation). If, on the other hand, participants generate both performed and observed actions, generation is not a feature diagnostic of source in this task. Yet, generation is typically diagnostic of internally-generated events, so participants may persist in evaluating generation to be indicative of self-performance. Critically, a diagnostic features account that assumes a dissociation of item and source memory would predict an effect of generation on source memory even in absence of an increase in item memory strength.

Two alternate accounts propose that memory judgements are not based on different qualitative features of memories, but rather on the overall memory strength of items.

Mistaken fluency account: This account proposes that participants recall the most fluent items when given a retrieval task, and mistake the inherent fluency of an item as a correct response given the retrieval task (Jacoby, Kelley, Brown, & Jasechko, 1989; Jacoby, Woloshyn, & Kelley, 1989; Marsh & Landau, 1995). If generating a performed actions overall leads to stronger memory for those performed actions, those actions should be more fluent and hence more likely to be reported at test. This account would then predict that generated and performed actions would be more likely to be reported than their non-generated counterparts regardless of the constraints of participants' recall (i.e., to recall performed or observed actions). This predicts more correct responses in the Recall-performed and more source errors in the Recall-observed task when performed actions were self-generated.

Relative strength account: This account assumes that the various qualitative features encoded for an item (and used for source decisions under a diagnostic features account) collapse into an overall memory strength representation of that item. Different sources of items are represented by distributions of different relative strengths akin to a one-dimensional model of signal detection (Hoffman, 1997; Marsh & Bower, 1993; Marsh & Landau, 1995). The relative strength account assumes heuristic responding at test based on the relative strength of items and the given retrieval task. Under this account, participants evaluate the overall memory strength of items. When asked to retrieve items from the stronger source, stronger items are reported and weaker items withheld. When asked to retrieve items from the weaker source, participants will now report weaker items and withhold stronger items. Critically, participants do not evaluate separate features of events at retrieval but only evaluate an event's overall strength. Under this account, if generating an action increases the memory strength of that action, those stronger actions are more easily identified as performed (the stronger source). This account predicts participants are going to recall more actions correctly as self-performed and fewer falsely as self-performed if they generated the idea for that action. Critically, the account also predicts that this increase in source memory performance is due to higher memory strength of performed actions after generation relative to memory strength of actions performed on instruction.

All three accounts make different predictions about the effects of generating the idea for an action on item and source memory. The diagnostic features account predicts that generating actions increases claims of ownership over those actions, without increasing item memory. The mistaken fluency account predicts that generating actions leads to an increase in fluency for those items, with generated actions more likely to report at test regardless of the retrieval orientation. The relative strength account predicts that source decisions are based on items' overall memory strength, with generating actions resulting in higher overall memory and thus increasing identification of generated actions as self-performed.

To test those predictions, I returned to the extended recall paradigm first introduced in Experiment 3 for all three experiments in this chapter. In the extended (selective) recall paradigm, participants take turns performing and observing actions at encoding. At retrieval, participants are asked to recall actions they performed or actions they observed. In contrast to the standard selective recall task, participants are asked to produce all actions that come to mind and then decide for each action if it complies with their recall task or does not. This way, this paradigm allows separate assessment of the overall availability of actions at retrieval as a measure of items' overall strength, and of participants' ability to monitor the source of an item at retrieval. Participants were asked to recall performed or observed actions to allow separate evaluation of the impact of generation under both retrieval task orientations. The action stimuli were the verb-objects interactions from Experiment 4. Unlike the intransient actions from Experiments 1 through 3, these lend themselves to be verbalized and generated as well as instructed for performance.

4.3 Experiment 5

Experiment 5 tests whether generating the idea for an action (to make which objects interact in what particular way) prior to performance improves memory for those performed actions, by either strengthening memory for those actions overall or making those actions more clearly diagnostic of having been self-performed, or both. Participants performed and observed actions. Half of the actions they performed they generated themselves, the other half they executed on verbal instruction from the experimenter. At test I compared memory performance between participant-generated and experimenter-provided sets of items.

4.3.1 Method

Participants

44 participants participated for partial course credit. 1 participant did not attend all sessions and their data were excluded from the analysis.

Procedure

Participants attended two sessions. Participants attended the first session in pairs. They sat facing one another across a table. Six sets of six objects each were set out a set at a time on the table between participants. The sets of objects were the same as in Experiment 4. For half the sets, participants were asked to generate verb-object phrases and perform them, as in Experiment 4. For the other half of sets, the experimenter read verb-object phrases to the participants and asked participants to perform the action.

All verb-object phrases described interactions of two objects, as in Experiment 4. Possible verb-object phrases were demonstrated to participants prior to the experiment using objects not used in the experiments, containing both static interactions (i.e., set object A next to object B) and dynamic interactions (i.e., throw object A at object B). For the sets with experimenter-provided actions, action phrases were created based on a pilot study to mimic typical patterns of actions generated by pairs of participants, including some overlap of combinations of objects.

Participants were asked to verbalize the verb-object phrases they self-generated prior to performing the actions. Participants heard the experimenter read the verb-object phrases out loud for the sets with experimenter-provided actions.

Participants took turns to perform actions for all of the sets. When self-generating actions, participants were reminded to avoid repeating an action (that is, a complete verb-objects phrase) that had already been performed by them or their partner. Each participant performed 10 actions for each set, resulting in each participants performing 30 participant-generated actions and 30 experimenter-provided actions.

Participants returned the next day individually for the retrieval session. The sets of objects were presented to them, one set at a time. The objects were arranged in a different pattern to the one that had been presented to participants at encoding to ensure that participants' responding was based on memory for the verb-objects interactions and not memory for location. Participants performed an extended recall task as in Experiment

3. Participants were randomly assigned to the Recall performed or Recall observed conditions in an extended recall task as in Experiment 3. Participants were, according to their condition, asked to re-perform actions they had performed at encoding or they had observed their partner perform, but to perform everything that came to mind as they tried to do the task. For each action they performed at retrieval, they were asked to indicate whether the action was task-compliant (i.e. one of their performed actions in the Recall performed, one of their partner's actions in the Recall observed task) or not. Actions were coded as matching actions from encoding only if the entirety of the action, i.e. both the objects and the way they interacted, was reproduced at retrieval.

4.3.2 Results

Encoding phase

I looked at two aspects of participants performance in the encoding phase. First, I looked at whether self-generated verb-object phrases matched experimenter-provided ones in terms of their typicality for each of the sets. Second, I looked at the sets with participant-generated actions in particular for repetition of self-generated and partner-generated items as in the prior experiments.

For the sets with experimenter-provided actions, verb-object phrases were read to participants, and participants were asked to merely execute the actions. For sets with participant-generated actions, participants were asked to generate a verb-object phrases themselves and execute it immediately after generation, with the constraint of using two objects at a time. To test whether verb-object phrases are roughly equivalent in both conditions, I tested whether they were equally typical verb-object phrases. If typicality of both diverges significantly, this could limit interpretation of the results.

I created frequency norms for the verb-object phrases generated by participants as I did for the actions Chapter 2. For each set of items, I summed up the number of times a verb-object phrase was generated by participants in the experiments and estimated the relative probability for an action phrase to be generated by dividing the number of times a phrase was generated by the total number of actions participants were asked to generate for that set (10 actions per participant, for the total number of participants). Experimenter-provided actions that were never generated by participants were added with the probability of half an occurrence to the frequency norms. The frequency norms for each set were then renormed so that probabilities in each set added to 1.

Table 4.1. Typicality of verb-object phrases in encoding phase in Experiment 5

Sets	Experimenter-provided	Self-generated	p
1	0.42 (0.12)	0.49 (0.13)	$p = .18$
2	0.49 (0.03)	0.49 (0.15)	$p = .95$
3	0.28 (0.09)	0.49 (0.08)	$p < .001$
4	0.45 (0.02)	0.47 (0.14)	$p = .51$
5	0.49 (0.09)	0.49 (0.10)	$p = .95$
6	0.39 (0.10)	0.50 (0.06)	$p = .003$

Note. SDs in brackets.

I looked at the frequency (or typicality) of verb-object phrases produced by participants to compare participant-generated and experimenter-provided actions. By performing a median-split on each set of items, I designated common items with 1 and rare items with 0, for a balanced average of common and rare items at 0.5. The average frequencies for both generation conditions by set are in Table 4.1.

While actions were of equal typicality for the majority of the sets, the actions generated by participants were more typical for two of the sets. Subsequent analyses of the data with both sets excluded revealed the same pattern of data as the analysis based on all sets, so the analyses reported are based on the full set of data. Possible implications of a difference in typicality by generation condition will be discussed later.

As in prior experiments, I examined participants' tendency to repeat items (that is verb-object phrases) they had already generated for a set themselves or that their partner had already generated. Compared to Experiment 4, encoding phase errors were quite frequent. Expressed as an average percentage of all actions generated by participants, self-repeats occurred in 2.42% (SD=2.53%) of performed actions and partner-repeats occurred in 5.23% (SD=2.59%) of performed actions. Partner-repeats were excluded from the subsequent analyses to ensure that each action was uniquely only performed by one person in the pair, as in previous experiments.

Test phase

Participants performed the extended recall task. This task allows me to look separately at all actions available to participants at retrieval (availability), the rates with which they report those actions as task-compliant (monitoring), and the frequency of actions they finally report as task-compliant. I will first look at participants' final report.

Table 4.2. Frequency of final responses in Experiment 5

	Generation	
Responses	Experimenter-provided	Participant-generated
Recall performed		
Correct responses	8.55 (2.48)	10.95 (3.36)
Source errors	1.82 (1.65)	2.23 (1.38)
Intrusion errors	4.36 (3.99)	3.86 (3.04)
Recall observed		
Correct responses	9.10 (2.84)	8.71 (3.32)
Source Errors	3.62 (2.62)	3.38 (2.31)
Intrusion Errors	6.51 (5.34)	6.19 (2.44)

Note. Mean with Standard Deviations in brackets.

Final report The mean performance in the final report, that is frequency of responses participants confirm as task-compliant in the extended recall task, is shown in Table 4.2. I will begin first by reporting the full analysis of the data, and then report the results of the planned comparisons (contrasting the memorial result of generating or executing actions on instruction).

I analysed the data with separate 2 Task (Recall performed, Recall observed) x 2 Generation (Sets with experimenter-provided actions, Sets with participant-generated actions) mixed ANOVAs with repeated measures on the second factor for correct responses, source errors and intrusion errors. To reiterate, Generation was blocked by sets. Participants self-generated actions for half the sets, and executed experimenter-provided actions for the other half. This means that for half the sets, self-performed actions were also self-generated, for the other half, they were experimenter-provided. Observed actions were partner-generated for half the sets, and experimenter-provided for the other half of sets.

For correct responses, there were no significant main effects of Task, $F(1,41) = 1.53$, $MSe = 10.06$, $p = .22$, $\eta^2_{\text{p}} = .036$, or Generation, $F(1,41) = 2.70, MSe = 8.20, p = .11, \eta^2_{\text{p}} = .062$, but there was a significant interaction, $F(1,41) = 5.10, MSe = 8.20, p = .029, \eta^2_{\text{p}} = .111$. I looked separately at the effects of Generation in the Recall performed and Recall observed task. Participants in the Recall performed task correctly reported more self-generated than experimenter-provided actions, $F(1,41) = 1.53, MSe = 10.06, p = .22, \eta^2_{\text{p}} = .036$. Observed actions were never self-generated, but generated by the partner or the experimenter. So unsurprisingly, there was no significant effect of Generation on correct retrieval in the Recall observed task, $F < 1$.

For source errors, there was a main effect of Task, $F(1,41) = 8.45, MSe = 5.55, p = .006, \eta_{\mathrm{p}}^2 = .171$. Source errors were more common in the Recall observed than the Recall performed task. There was no significant main effect of Generation or interaction, both $F < 1$. Intrusion errors occurred more frequently in the Recall observed than Recall performed task, $F(1,41) = 4.87, MSe = 22.22, p = .033, \eta_{\mathrm{p}}^2 = .106$, but there was no significant main effect of Generation or a significant interaction either, both $F < 1$.

According to a diagnostic features and a relative strength account, generating actions should increase their retrieval as correct responses in the Recall performed task and decrease source errors in the Recall observed task. The fluency account on the other hand would predict an increase in source errors in the Recall observed task after generation. While the significant interaction for correct responses showed the predicted effect, the interaction was not significant for source errors. I looked specifically at source errors in the Recall observed task to examine the predicted effect of Generation as outlined above. While self-generated self-performed actions were less likely to be falsely retrieved as observed than experimenter-provided actions, this difference was not significant, $t(20) = .41, p = .34$, one-tailed. This null-effect does not allow distinguishing the theoretical accounts.

Participants' responses in the final report are the product of the number of performed, observed and novel actions available to participants at retrieval and the rates with which these actions are accepted as task-compliant. Thus the effect of Generation observed in the final report data could be the result of performed actions coming to mind more frequently at retrieval overall (for example due to higher memory strength), performed actions being more likely to be identified as performed at monitoring due to specific source features, or a combination of the two.

Availability The extended recall task allows me to separately look at the number of actions available to participants at retrieval and the monitoring choices they make about those actions. I looked at all actions participants generated at retrieval first, irrespective of participants' later choice to withhold some of those actions from report.

I first looked at the overall number of actions participants produced at retrieval to test if one retrieval task (Recall performed or Recall observed) or initial encoding instruction (participant-generated sets, experimenter-generated sets) is more likely to make participants perform actions at retrieval. Participants in the Recall performed task pro-

Table 4.3. Frequency of actions generated at retrieval in Experiment 5

	Generation	
Source	Experimenter-provided	Participant-generated
Recall performed		
Performed actions	10.59 (3.13)	11.82 (3.30)
Observed actions	7.45 (3.71)	7.55 (2.50)
Novel actions	8.68 (10.01)	7.05 (5.63)
Recall observed		
Performed actions	9.14 (4.28)	9.76 (5.75)
Observed actions	10.05 (2.85)	9.57 (3.79)
Novel actions	9.05 (7.12)	8.24 (3.35)

Note. Mean with Standard Deviations in brackets.

duced 26.73 actions (SD=11.89) for sets with experimenter-provided actions and 26.41 (SD=7.61) in response to sets with participant-generated actions. In the Recall observed task, participants produced on average 28.24 (SD=10.30) actions in response to sets with experimenter-provided actions and 27.58 (SD=10.76) in response to sets with self-generated actions. Overall a 2 Source (Recall performed, Recall observed) x 2 Generation (experimenter-provided sets, participant-generated sets) mixed ANOVA with repeated measures on the second factor on the total number of actions produced by participants at retrieval showed no evidence for a significant effect of Task or Generation, nor was there a significant interaction, all $F < 1$.

I next looked at the source of the actions participants produced at retrieval, prior to making a judgment about the retrieved action complying with their retrieval task. The mean frequencies for this part of the test phase are presented in Table 4.3. I first analysed actions generated at retrieval with a 2 Task (Recall performed, Recall observed) x 2 Source (Performed, Observed) x 2 Generation (Sets with experimenter-provided actions, Sets with participant-generated actions) mixed ANOVA with repeated measured on the second and third factor.

There were no significant main effects of Task or Generation, $Fs < 1$, but there was a significant main effect of Source, $F(1,41) = 9.95, MSe = 12.10, p < .001, \eta_p^2 = .195$. Performed actions were generated more frequently at retrieval than observed actions. The effect of Source was moderated by a significant interaction with Task, $F(1,41) = 14.65, MSe = 12.10, p < .001, \eta_p^2 = .263$. While performed actions were generated more frequently than observed actions in the Recall performed task, $F(1,21) = 44.08, MSe = 6.85, p < .001, \eta_p^2 = .677$, there was no evidence for performed actions being generated more frequently than

Table 4.4. Rates of actions generated at retrieval that were reported as task-compliant in Experiment 5

	Generation	
Source	Experimenter-generated	Self-generated
Recall performed		
Performed actions	0.82 (0.14)	0.92 (0.10)
Observed actions	0.24 (0.20)	0.30 (0.18)
Novel actions	0.48 (0.27)	0.59 (0.28)
Recall observed		
Performed actions	0.45 (0.29)	0.42 (0.29)
Observed actions	0.91 (0.08)	0.92 (0.08)
Novel actions	0.70 (0.28)	0.77 (0.16)

Note. Mean with Standard Deviations in brackets.

observed actions in the Recall observed task, $F < 1$. The remaining interactions were not significant, all $F < 1.83, p > .18$. This suggests that the effects of Generation in the final report are not simply an effect of self-generated actions being stronger than experimenter-generated actions.

I separately looked at participants' propensity to generate novel actions at retrieval. There were no significant main effects of Task or Generation, nor was there a significant interaction, all $F < 2.09, p > .15$. This suggests that neither the retrieval task or generation condition drove responding.

Monitoring I finally analysed the rates with which actions generated at retrieval were reported as complying with the retrieval task. The means for this phase of the retrieval task are in Table 4.4. I analysed the rates with which encoded actions were accepted as task-compliant with a 2 Task (Recall performed, Recall observed) x 2 Source (Performed, Observed) x 2 Generation (Sets with experimenter-provided actions, Sets with participant-generated actions) mixed ANOVA with repeated measured on the second and third factor.

There were significant main effects of Task, $F(1,41) = 8.99, MSe = 0.05, p = .004, \eta_p^2 = .180$, and Source, $F(1,41) = 4.09, MSe = 0.04, p = .050, \eta_p^2 = .091$. Overall, participants in the Recall observed task reported a higher rate of available actions as task-compliant, and overall performed actions were more likely to be reported as task-compliant than observed actions. There was no significant main effect of Generation, $F(1,41) = 2.86$, $MSe = 0.02, p = .098, \eta_p^2 = .065$. The main effects were moderated by significant Task by Source interaction, $F(1,41) = 345.04, MSe = 0.04, p < .001, \eta_p^2 = .894$. Unsurprisingly, performed actions were reported at higher rates than observed actions in the Recall performed task, $F(1,21) = 324.36, MSe = 0.02, p < .001, \eta_{\mathrm{p}}^2 = .939$, while the reverse was true in the Recall observed task, $F(1,21) = 99.39, MSe = 0.05, p < .001, \eta_{\mathrm{p}}^2 = .832$. The Task by Generation interaction was marginally significant, $F(1,41) = 3.81, MSe = 0.02, p = .058, \eta_{\mathrm{p}}^2 = .085$. In the Recall performed task, performed and observed actions retrieved for sets with participant-generated actions were accepted as task-compliant at higher rates ($M_{SG} = 0.61$, $SD_{SG} = 0.34$) than actions retrieved in response to experimenter-provided sets ($M_{EG} = 0.53$, $SD_{EG} = 0.34$), $F(1,21) = 10.39, MSe = 0.01, p = .004, \eta_{\mathrm{p}}^2 = .331$. In the Recall observed task, there was no evidence for a difference in rates between sets with participant-generated actions ($M_{SG} = 0.67, SD_{EG} = 0.33$) and sets with experimenter-provided actions ($M_{EG} = 0.68, SD_{EG} = 0.31$), $F < 1$. Neither the Generation by Source or critically the three-way interaction reached significance, both $F < 1$.

As before, I had clear predictions about the effects of Generation in this part of the retrieval task. Under a diagnostic features that predicts systematic evaluation of source features, performed actions that were self-generated should be identified as self-performed at higher rates than experimenter-provided actions. It follows that in the Recall performed task, self-generated performed actions should be accepted as task-compliant at higher rates, and that was indeed the case, $F(1,21) = 9.97, MSe = 0.01, p = .005, \eta_{\mathrm{p}}^2 = .322$. In the Recall observed task, I predicted that self-generated performed actions should be more likely to be withheld from report than experimenter-provided performed actions, but there was no evidence for this, $F < 1$.

Finally I looked at the rates with which novel actions were reported as task-compliant. Not all participants generated novel actions for both types of set at retrieval, so they provided no evidence for monitoring of the novel actions. Rather than excluding them entirely, I analysed the data with the linear mixed model analogue of the repeated measures ANOVA using the *afex* package (Singmann, Bolker, & Westfall, 2015) where participants are included in the model as random factors. There was a significant main effect of Task for the rates of accepting novel actions as task-compliant, $F(1,40.76) = 13.58, p < .001$, with higher rates of acceptance in the Recall observed than Recall performed task. There was no significant effect of Generation, $F(1,40.76) = 2.62, p = .11$, nor was there a significant interaction, $F < 1$.

4.3.3 Discussion

In Experiment 5, I tested whether generating actions affects subsequent memory of self-performed actions (compared to only performing actions on instruction). Predictions of various source memory accounts were based on possible dissociations of overall item memory strength and specific evaluation of source memory features.

One concern with giving participants the freedom to generate their own actions is that they will be fundamentally different from experimenter-provided actions. I did find that experimenter-provided ideas were on average less typical than participant-generated ideas for two of the sets of objects used in the experiments. If that typicality translates into fluency, any additional strength of generated actions may be due to the inherent fluency of the ideas rather than the cognitive processing involved in generating ideas. Running the analyses without the two sets with discrepant typicalities led to identical results, so any effects of Generation are not merely a reflection of typicality-based fluency effects. This provides some evidence against a mistaken fluency account that would have predicted that higher typicality (and hence higher fluency of those actions) would be critically responsible for the pattern of responding.

The experiment showed a strong overall enactment effect, with performed actions more available than observed actions at retrieval. However, this enactment effect was moderated by the orientation towards the target, with performed actions being not significantly more likely to be available at test than observed actions, if participants attempt to retrieve observed actions. This suggests that a focus on the weaker source of items can limit the benefit of encoding actions by enactment.

Typically, generating verbal items increases the overall memory strength of those items (Slamecka & Graf, 1978). In the extended recall paradigm, higher overall memory strength should be reflected in a higher availability of actions regardless of the retrieval task orientation under the fluency, relative strength and diagnostic features account. While there were numeric trends for more performed actions being available to participants when they were generated rather than experimenter-provided, this was not statistically significant. This means there was no evidence for a generation effect for actions. Similar results have been reported by Nilsson and Cohen (1988) and Lichty, Bressie, and Krell (1988). In a fully crossed between-subjects paradigm, Nilsson and Cohen asked participants either to generate action phrases themselves or provided them with action phrases, and asked participants to perform them or merely study them without performance. While both performing actions and generating actions increased later recall, enactment and generation did not combine to further increase recall than either encoding strategy on their own. Lichty et al. asked participants to perform provided action phrases or generate and subsequently perform them. They did not observe a significant benefit of generation on memory recall for performed phrases. Theoretical accounts of enactment argue that performing an action (much like generating a verbal item) increases item-specific processing of the item (Engelkamp & Dehn, 2000; McDaniel & Bugg, 2008; Steffens, 2007). Additional enrichment by deeper level of processing does not further increase the overall memory strength of an item (Nilsson & Cohen, 1988; Nilsson et al., 1995; Zimmer & Engelkamp, 1999).

The second critical aspect of the data was participants' ability to correctly report and withhold items as a measure of their source memory performance. In fact, when participants had generated the ideas for actions they performed, they were more likely to correctly report those actions as performed. Under a fluency account, the higher memory strength of self-generated self-performed actions should should have led to an increase in false reports of those actions as observed, while a relative strength account would predict that participants would be more likely to be able to withhold those items from report. The diagnostic features account made similar quantitative predictions to the relative strength account, but argues that it is not the availability of memory strength of items that drives participants' ability to withhold them from report, but the cognitive features characteristic of performed actions. Yet, there was no significant evidence for participants' showing an increased ability to withhold those actions when recalling actions they observed their partner perform.

What does that mean for the possible accounts of source memory performance? There was no clear increase in overall memory strength for performed actions after generation, suggesting that memory strength alone could not have accounted for the present pattern of data. A source monitoring account on the other hand would argue that participants evaluate specific features of items. For performed actions that have been self-generated, participants could use that knowledge of self-generation as an additional indicator of the action being internally-generated (i.e., performed). It is that specific qualitative trace that led participants to correctly report higher rates of performed actions in the Recall performed task. However, there was no evidence for participants using that same knowledge to withhold performed actions from report. The predicted interaction failed to reach significance.

Generating actions participants performed did not improve their memory retrieval for those actions, compared to performing actions on instruction. While generating actions allowed participants to better identify performed actions as performed, it did not prevent them from falsely reporting performed actions as observed. In other words, the effects of generation on memory for actions are not very convincing. It is possible that participants were aable to use alternate perceptual features to make correct source decisions. Participants sat facing one another across the table. The visual trace associated with performed and observed actions would give a clear indicator of an action having been performed or observed, without participants having to consider the cognitive trace associated with performed actions. Secondly, the arrangement of the objects in the sets varied between encoding and retrieval to ensure that participants could not rely on memory of location but had to remember the exact verb-objects phrases. While the exact action necessary for interaction of two objects would still be identical, the movement of choosing objects from the set arrangement would differ between both encoding phases. It is possible that those spatial cues were crucial in the generation decisions participants made for those sets. Without having the same visual cues available to them at retrieval, participants may not have been able to reactivate the same generated processes when retrieving actions at test. In Experiment 6 those alternate cues will be minimized to strengthen a possible effect of the generation manipulation.

4.4 Experiment 6

In Experiment 6, I kept the arrangement of sets of objects constant between encoding and retrieval. I additionally boosted the relative importance of the cognitive trace of having generated an action, by increasing the visual similarity of performed and observed actions. Participants observed their partner's perform actions from a first-person perspective. I additionally changed some of the experimenter-provided actions to more closely mirror participant-generated actions.

Table 4.5. Typicality of verb-object phrases in encoding phase in Experiment 6

Sets	Experimenter-provided	Self-generated	p
1	0.39 (0.02)	0.50 (0.11)	$p = .009$
2	0.40 (0.00)	0.48 (0.10)	$p = .022$
3	0.45 (0.00)	0.49 (0.10)	$p = .20$
4	0.56 (0.01)	0.50 (0.09)	$p = .51$
5	0.36 (0.02)	0.48 (0.11)	$p = .003$
6	0.37 (0.03)	0.49 (0.13)	$p = .013$

Note. SDs in brackets.

4.4.1 Method

Participants

46 participants participated for partial course credit. Data from two participants was excluded because they did not follow the instructions at retrieval.

Procedure

The procedure, action objects and retrieval tasks were identical to Experiment 5 except for the following changes. The sets of objects were arranged in identical arrangement at both encoding and retrieval, and identical for both participants. As in the previous experiment, participants took turns performing actions. In this experiment, participants observed their partner's action performance by watching over their partner's shoulder. Participants switched from performing to observing and thus sitting down to standing up after every action. The actions were verbalized prior to performance by the participant performing the action in the participant-generated sets and by the experimenter in the experimenter-provided sets. Finally, some of the actions from the experimenter-provided sets in Experiment 5 had never been performed by participants, so those actions were exchanged for low-frequency actions that participants had in fact generated to even out the typicality of participant-generated and experimenter-provided actions.

4.4.2 Results

Encoding phase

As in Experiment 5 I first looked at the typicality of actions provided by the experimenter and generated by participants. The average frequencies for both conditions by set are in Table 4.5. Items in experimenter-generated sets were on average less typical than items in participant-generated sets.

Second, I examined participants' tendency to repeat items they had already generated for a set themselves or that their partner had already generated. Expressed as an average percentage of all actions generated by participants, self-repeats occurred in 0.80% (SD=1.60%) of performed actions and partner-repeats occurred in 2.03% (SD=1.81%) of performed actions. Partner-repeats were excluded as before.

Test phase

As in Experiment 5, participants completed an extended recall. I will first discuss participants' final report, and then separately discuss the two extended recall phases.

Final report The mean frequencies for responses participants reported at retrieval are shown in Table 4.6.

I analysed participants' reported responses with multiple 2 Task (Recall performed, Recall observed) x 2 Generation (Sets with experimenter-provided actions, Sets with participant-provided actions) repeated measures ANOVAs with repeated measures on the second factor separately for correct responses, source errors and intrusion errors. For correct responses, there was no significant main effect of Task, $F(1,42) = 2.72, MSe = 9.63, p = .11, \eta_{\text{p}}^2 = .061$. There was a significant main effect of Generation, $F(1,42) = 5.05, MSe = 7.06, p = .030, \eta_{\text{p}}^2 = .107$. Participants were overall more likely to report correct responses when they or their partner had generated the ideas for actions. This effect was qualified by a Task by Generation interaction, $F(1,42) = 4.35, MSe = 7.06, p = .043, \eta_{\text{p}}^2 = .094$. Participants in the Recall performed task correctly recalled more self-generated than experimenter-provided actions, $F(1,21) = 9.82, MSe = 6.75, p = .005, \eta_{\text{p}}^2 = .319$. In the Recall observed task, there was no evidence that correct recall was higher in the sets with participant-generated actions than for sets with experimenter-provided actions, $F < 1$.

For source errors, there was a main effect of Task, $F(1,42) = 9.75$, $MSe = 3.52$, $p = .003$, $\eta_{\text{p}}^2 = .188$. Source errors were common in the Recall observed than the Recall performed task. There was a significant main effect of Generation, $F(1,42) = 11.62, MSe = 3.18, p = .001, \eta_{\text{p}}^2 = .217$. This suggests that source errors were committed more frequently when participants retrieved actions from sets with experimenter-provided actions than from sets with participant-generated actions. The interaction of Task and Generation was not significant, $F(1,42) = 1.29, MSe = 3.18, p = .26, \eta_{\text{p}}^2 = .030$.

There was no significant main effect of Task on intrusion errors, $F(1,42) = 2.28, MSe =$

Table 4.6. Frequency of final responses in Experiment 6

	Generation	
Responses	Experimenter-provided	Participant-generated
Recall performed		
Correct responses	8.55 (3.07)	11.00 (2.67)
Source errors	2.77 (2.07)	1.91 (1.60)
Intrusion errors	3.68 (2.28)	4.73 (2.96)
Recall observed		
Correct responses	8.64 (2.72)	8.73 (3.07)
Source Errors	4.45 (1.92)	2.73 (1.70)
Intrusion Errors	5.05 (4.15)	6.23 (4.69)

Note. Mean with Standard Deviations in brackets.

$19.78, p = .14, \eta_p^2 = .051$. There was a marginally significant main effect of Generation, $F(1,42) = 4.02, MSe = 6.79, p = .051, \eta_p^2 = .087$. This suggests that novel actions were more frequently part of participants' report when retrieving actions from participant-generated than experimenter-provided sets. There was no significant interaction, $F < 1$.

Availability The extended recall task allows me to look at the two phases of participants' recall separately that result in the responses participants report as task-compliant responses.

As in Experiment 5, I first looked at the overall number of actions participants produced at retrieval prior to reporting or withholding the item to ascertain if either retrieval task or experimental manipulation would make participants more willing to produce items at test. I tested the total number of actions produced by participants with a 2 Task (Recall performed, Recall observed) x 2 Generation (Sets with experimenter-provided actions, Sets with participant-generated actions) ANOVA with repeated measures on the second factor. Overall there was no evidence for significant main effects of Task, $F(1,42) = 1.93, MSe = 64.86, p = .17, \eta_p^2 = .044$, or Generation, $F < 1$. However, there was a significant interaction, $F(1,42) = 4.72, MSe = 9.55, p = .035, \eta_p^2 = .101$. Participants in the Recall performed task generated more actions at retrieval in response to sets with participant-generated actions ($M_{PG} = 28.18, SD_{PG} = 5.28$) than in response to sets with experimenter-provided actions ($M_{EP} = 26.13, SD_{EP} = 5.28$), $F(1,21) = 5.81, MSe = 7.93, p = .025, \eta_p^2 = .217$. In the Recall observed task, there was no evidence that more actions were generated in response to sets with participant-generated actions ($M_{PG} = 29.95, SD_{PG} = 7.81$) than in response to sets with experimenter-provided actions ($M_{EP} = 29.14, SD_{EP} = 5.67$), $F < 1$.

Table 4.7. Frequency of actions generated at retrieval in Experiment 6

	Generation	
Source	Experimenter-provided	Participant-generated
Recall performed		
Performed actions	10.86 (3.11)	12.50 (3.00)
Observed actions	8.77 (2.47)	7.59 (2.70)
Novel actions	6.50 (3.92)	8.09 (3.88)
Recall observed		
Performed actions	11.73 (3.06)	10.50 (3.02)
Observed actions	10.68 (2.63)	9.68 (3.15)
Novel actions	7.55 (6.27)	8.95 (5.74)

Note. Mean with Standard Deviations in brackets.

I next looked at the kind of actions participants produced at test by looking at the frequency of performed, observed and novel actions generated by participants. The means for this phase of the retrieval are presented in Table 4.7. I analysed this phase of the retrieval task with a 2 Task (Recall performed, Recall observed) x 2 Source (Performed, Observed) x 2 Generation (Sets with experimenter-provided actions,Sets with participant-generated actions) mixed ANOVA with repeated measured on the second and third factor.

There were no significant main effects of Task, $F(1,42) = 1.75, MSe = 12.90, p = .19, \eta_p^2 = .040$ and Generation, $F(1,42) = 1.51, MSe = 5.71, p = .22, \eta_p^2 = .035$. There was a marginally significant Task by Generation interaction, $F(1,42) = 3.46, MSe = 5.72, p = .070, \eta_p^2 = .076$. There was a significant main effect of Source, $F(1,42) = 32.56, MSe = 6.63, p < .001, \eta_p^2 = .437$, with performed actions overall retrieved more frequently than observed actions. This was qualified by a significant Task by Source interaction, $F(1,42) = 10.94, MSe = 6.63, p = .002, \eta_p^2 = .207$. Performed actions were retrieved more frequently than observed actions in the Recall performed task, $F(1,21) = 42.08, MSe = 6.40, p < .001, \eta_p^2 = .667$. This effect was not significant in the Recall observed task, $F(1,21) = 2.78, MSe = 6.86, p = .11, \eta_p^2 = .117$. The Generation by Source interaction was not significant, $F(1,42) = 2.19, MSe = 8.44, p = .15, \eta_p^2 = .050$, but the three-way interaction was marginally significant, $F(1,42) = 3.02, MSe = 8.44, p = .089, \eta_p^2 = .067$.

I specifically looked at the effect of generating performed actions on the availability of those actions at retrieval. In the critical 2 Task (Recall performed, Recall observed) x 2 Generation (Sets with experimenter-provided actions, Sets with participant-generated actions) ANOVA with repeated measures on the second factor, I predict a main effect of Generation if generating performed actions increases their memory strength. In fact,

Table 4.8. Rates of actions generated at retrieval that were reported as task-compliant in Experiment 6

	Generation	
Source	Experimenter-provided	Participant-generated
Recall performed		
Performed actions	0.78 (0.14)	0.88 (0.13)
Observed actions	0.30 (0.18)	0.25 (0.21)
Novel actions	0.59 (0.22)	0.58 (0.18)
Recall observed		
Performed actions	0.39 (0.17)	0.27 (0.14)
Observed actions	0.80 (0.13)	0.89 (0.14)
Novel actions	0.68 (0.20)	0.70 (0.18)

Note. Mean with Standard Deviations in brackets.

there were no significant main effects of Task or Generation, both $F < 1$, but there was a significant interaction, $F(1,42) = 6.26, MSe = 7.20, p = .016, \eta_p^2 = .130$. Performance in the Recall performed task shows a marginally significant effect in the predicted direction, $F(1,21) = 3.21, MSe = 9.17, p = .087, \eta_p^2 = .133$, while performance in the Recall observed task shows a marginally significant effect in opposite direction, $F(1,21) = 3.16, MSe = 5.23, p = .090, \eta_p^2 = .131$. This suggests that overall there is little evidence suggesting generation adds an overall increase in memory strength for performed actions.

I finally looked separately at the number of novel actions performed at retrieval. There was no significant main effect of Task, $F < 1$, but there was a marginally significant main effect of Generation, $F(1,42) = 3.92, MSe = 12.63, p = .054, \eta_p^2 = .085$. This indicates a trend for novel actions to be performed more frequently when retrieving actions in response to participant-generated sets than to experimenter-provided sets. The interaction was not significant, $F < 1$.

Monitoring I will now look at the rates with which participants judged actions available to them at retrieval to be task-compliant, i.e., actions they had performed themselves in the Recall performed task and actions they had observed their partner perform in the Recall observed task.

The means for this phase of the retrieval are presented in Table 4.8. I first analysed the rates with which encoded actions were reported as task-compliant with a 2 Task (Recall performed, Recall observed) x 2 Source (Performed, Observed) x 2 Generation (Sets with experimenter-provided actions, Sets with participants-generated actions) mixed ANOVA with repeated measured on the second and third factor.

There were no main effects of Task, $F(1,42) = 2.17, MSe = 0.02, p = .15, \eta_\text{p}^2 = .049$, Source and Generation, both $F < 1$. There were no significant Task by Generation, $F(1,42) = 1.05, MSe = 0.02, p = .31, \eta_\text{p}^2 = .024$, or Source by Generation interactions, $F(1,42) = 1.02$, $MSe = 0.02$, $p = .32$, $\eta_\text{p}^2 = .023$. Unsurprisingly, there was a significant Task by Source interaction, $F(1,42) = 379.69, MSe = 0.03, p < .001, \eta_\text{p}^2 = .900$. In the Recall performed task, participants reported performed actions at higher rates than observed actions as task-compliant, $F(1,21) = 208.33, MSe = 0.03, p < .001, \eta_\text{p}^2 = .908$, while the reverse was true in the Recall observed task, $F(1,21) = 172.57, MSe = 0.03, p < .001, \eta_\text{p}^2 = .892$. The results were qualified by the significant critical three-way interaction between Task, Source and Generation, $F(1,42) = 23.63, MSe = 0.02, p < .001, \eta_\text{p}^2 = .360$.

I looked at the Source by Generation interaction separately for the Recall performed and Recall observed task. The interaction was significant in the Recall performed task, $F(1,21) = 7.26, MSe = 0.02, p = .014, \eta_\text{p}^2 = .257$. Closer inspection showed that higher rates of performed actions were reported when actions had been self-generated rather than experimenter-provided, $F(1,21) = 9.85, MSe = 0.01, p = .005, \eta_\text{p}^2 = .319$. There was no evidence for an effect of Generation on the rates of falsely reported observed actions, $F < 1$.

The Source by Generation interaction was also significant in the Recall observed task, $F(1,21) = 17.62, MSe = 0.02, p < .001, \eta_\text{p}^2 = .456$. Here performed actions were falsely reported as observed with a lower rate when actions had been self-generated rather than experimenter-provided, $F(1,21) = 9.60, MSe = 0.02, p = .005, \eta_\text{p}^2 = .314$. At the same time, observed actions were more likely to be correctly reported as task-compliant if the actions had been generated by the partner rather than provided by the experimenter, $F(1,21) = 110.07, MSe = 0.01, p = .005, \eta_\text{p}^2 = .324$.

Finally I looked at participants' propensity to classify novel actions as task-compliant. Not all participants generated novel actions at retrieval, so I analysed the data with the mixed model approach using the *afex* package (Singmann et al., 2015) as in Experiment 5. There was a marginally significant main effect of Task, $F(1,41.75) = 3.97, p = .053$, with a trend to report novel actions more frequently as task-compliant in the Recall observed than Recall performed task. There was no significant main effect of Generation, nor was there an interaction, both $F < 1$.

4.4.3 Discussion

The pattern of data in Experiment 6 largely mirrored the pattern of data in Experiment 5, with the effects of the generation manipulation clearer than in the previous experiment. This suggests that maximizing the visual-spatial overlap between performed and observed actions did result in greater focus on the cognitive information during source decisions. I will return to look at the effects of visual perspective in Experiment 8 in Chapter 5. While the enactment effect was stronger than in the previous experiment, it was similarly moderated by the the retrieval task. When participants aimed to retrieve observed actions, the higher retrieval of observed actions relative to performed actions eradicated the enactment benefit.

Despite generating actions being relatively more crucial in making source decisions in the absence of the visual cue, there was no evidence that generating actions for self-performance increases the memory for those actions relative to actions performed on instruction. Yet there was clear evidence (compared to the partial evidence in Experiment 5) that generating actions for performance increases participants' ability to make correct source judgements about those actions. What this suggests therefore is that generation does not increase memory for performed actions, but that it increases participants' source memory, leading to a dissociation of item and source memory. This seems to provide more evidence in favour of a source monitoring account of source memory that argues for source memory decisions being made after evaluation of qualitative features of memories than accounts that argue that overall memory strength could account for the results.

One alternate possibility is that participants had two avenues to test for retrieved items in participant-generated sets: (1) Do I feel I performed this? Here participants would examine whether the motor information at retrieval matches self-performance of the action at encoding. (2) Do I feel I would have generated this? Here participants would examine whether it is plausible they would have come up with this particular idea for an action. In participant-generated sets, participants could test these two aspects of their memories to identify whether an action had been self-performed. In experimenter-provided sets, participants could only test the first question (Do I feel I performed this?) to identify an action as self-performed. Under this explanation, generating actions simply adds an additional cue for identification to one set of actions, regardless of whether that cue is diagnostic of the type of event (which would be the explanation favoured by a diagnostic features account). Experiment 7 will test whether this additional cueing is the critical mechanism for improved source memory performance, or if a stronger cognitive trace is specifically associated with performed actions.

There is another more trivial explanation for the source memory enhancement observed in this experiment. While participants generated and verbalized an action phrase themselves before they performed it in the sets with participant-generated actions, the experimenter verbalized all action phrases for both participants for the other sets. Therefore the verbalization (tone of voice, syntax) offered a cue diagnostic of source in one set of sets but failed to do so in the other. It is possible that this cue alone drives what looks like a generation effect in source memory in those two experiments by providing a clear distinction between self and partner in sets with participants-generated but not sets with experimenter-provided actions.

Beyond that it is possible that the types of actions provided to participants in both types of sets differed. While a brute-force analysis of typicality seems to show, at least, that the items that were generated and provided differ in this experiment, it is not clear what effect in particular that would have had on memory performance. If items in experimenter-provided sets were more typical and common overall, an increase in confusion would be plausible. Instead, items in experimenter-provided sets appeared to have a tendency to be less common. Hence, if participants were asked to perform them, and they remembered the item at all, this item should be less available at retrieval and if available, less prone to be source-misattributed (Engelkamp, Zimmer, & Biegelmann, 1993). There was no evidence for that. It is possible that items in experimenter-provided and participant-generated sets differed on dimensions not captured by a brute-forced typicality analysis (clustering along objects or actions, particular actions used, clustering of actions or objects by participant). While I treated them as equivalent in Experiment 5 and Experiment 6, and they appear equivalent on the surface, it is possible that this is not true. In Experiment 7, I therefore asked participants to take turns generating actions. Participants thus performed some self-generated and some partner-generated actions, with actions never provided by the experimenter.

This change in paradigm also addresses another possible reason for the absence of a generation effect on item memory. Generation of actions was blocked across sets, with assignment of encoding condition to sets counterbalanced across participants. Participants were presented with sets of objects one at a time and asked to either generate and perform actions or merely execute them on instruction of the experimenter. The generation effect as reported in the literature is stronger for intermixed lists of generated and non-generated items as a function of contrasting memory traces (Bertsch, Pesta, Wiscott, & McDaniel, 2007), in line with theoretical accounts that base the generation effect in inter-item distinctiveness (Hunt, 2012). In Experiment 7, Generation was therefore manipulated within-list.

4.5 Experiment 7

Previously, I found no evidence for a generation effect for the item memory of actions in two experiments. In Experiment 7, the critical comparison of generated versus not-generated actions was conducted within-sets when it had been between-sets in the previous experiments and between-subjects in Nilsson and Cohen (1988). If generation effects arise because generated items are preferentially retrieved in direct comparison to non-generated items, within-set manipulations of generation should boost the generation effect on item memory.

Secondly, participants will take turns generating actions for self- and partner-performance in Experiment 7. This eliminates the voice-cue being diagnostic for source in participant-generated but not experimenter-provided sets and eliminates potential differences between participant-generated and experimenter-provided actions. What follows is that generation per se is now no longer diagnostic of either source. If participants compare the profile of a given event with the prototypical source profile of that events (Mitchell & Johnson, 2000), participants should be more likely to claim self-generated actions as self-performed.

4.5.1 Method

Participants

46 members of the public participated for compensation of £8. Data from 2 participants was excluded because they did not attend all sessions.

Procedure

The objects and arrangement of sets to objects used was identical to Experiment 6. For each set of objects participants took turns generating actions out loud. After a participant generated an action, one of the participants would receive the instruction on an iPad to perform the action. While participants strictly took turns generating actions, the order of performance was randomized to ensure that participants could not predict whether it would be them or their partner performing any given action that was generated.

Participants generated 10 actions each for each set of objects, of which they performed 5 themselves and observed the remaining 5 be performed by their partner. In turn each participant heard their partner generate 10 actions and performed 5 of those and observed their partner perform the remaining 5 actions. This resulted in each participants generating 60 actions across all sets, of which they performed 30 actions themselves, and performing 60 actions across all sets, of which they had generated 30 actions themselves.

Participants returned the following day individually for the test phase and performed an extended recall task identical to Experiment 5 and Experiment 6. Critically, as in those experiments, participants were asked to recall actions they had performed or actions they had observed, regardless of who had generated the idea for the action.

4.5.2 Results

Encoding phase

Participants were instructed not to repeat actions they had generated themselves or that their partner generated. Expressed as a proportion of all actions generated, self-repeats occurred for 2.61% (SD=3.21%) and partner-repeats occurred for 4.64% (SD=3.13%) of cases. Partner-repeats were excluded as before.

Test phase

Participants performed an extended recall task at retrieval. As in previous experiments, I will first discuss participants' final report before separately looking availability and monitoring. The means for this phase of the retrieval task are in Table 4.9.

I analysed the data with multiple 2 Task (Recall performed, Recall observed) x 2 Generation (Partner-generated, Self-generated) mixed ANOVAs with repeated measures on the second factor, separately for correct responses and source errors. For correct responses, there was no significant main effect of Task, $F < 1$. There was a significant main effect of Generation, $F(1,42) = 6.57, MSe = 6.65, p = .014, \eta_p^2 = .135$, suggesting that overall participants correctly recalled self-generated actions more frequently than partner-generated actions. The effect was qualified by a marginally significant Task by

Table 4.9. Frequency of final responses in Experiment 7

	Generation	
Responses	Partner-generated	Self-generated
Recall performed		
Correct responses	8.36 (3.05)	8.77 (4.12)
Source errors	4.00 (3.02)	4.05 (2.44)
Intrusion errors	13.27 (8.01)	
Recall observed		
Correct	7.59 (3.22)	10.00 (4.14)
Source Errors	4.50 (3.19)	5.60 (2.50)
Intrusion Errors	12.73 (6.06)	

Note. Mean with Standard Deviations in brackets.

Generation interaction, $F(1,42) = 3.31, MSe = 6.65, p = .076, \eta_p^2 = .073$. There was no evidence for self-generated actions being correctly reported more frequently than partner-generated actions in the Recall performed task $F < 1$, while participants in the Recall observed task correctly retrieved more observed actions when they had generated the actions themselves, $F(1,21) = 10.42, MSe = 6.13, p = .004, \eta_p^2 = .331$. There were no significant main effects or significant interaction for source errors participants reported, all $F < 1.96, p > .16$. Since generation was manipulated within-set, intrusion errors could only be analysed across Generation conditions. There was no evidence that more intrusion errors were reported as task-compliant when recalling observed than when recalling performed actions, $F < 1$.

Availability In the extended recall task, participants were asked to perform all actions that came to mind as they tried to retrieve performed or observed actions before deciding for each action whether it complied with their recall task. I first looked at the total number of actions participants performed at retrieval in both retrieval tasks. There was no evidence that participants performed more actions at test in the Recall performed (M=68.00, SD=21.68) than the Recall observed task (M=66.95, SD=14.03), $F < 1$.

I next looked the frequency of performed, observed and novel actions available to participants at retrieval. The means for this phase of the retrieval task are in Table 4.10.

I analysed the data with a 2 Task (Recall performed, Recall observed) x 2 Generation (Partner-generated, Self-generated) x 2 Source (Performed, Observed) mixed ANOVA with repeated measures on the second and third factor. There was no significant main effect of Task, $F < 1$. There was a strong main effect of Generation, $F(1,42) = 18.72, MSe =$

Table 4.10. Frequency of actions generated at retrieval in Experiment 7

	Generation	
Source	Partner-generated	Self-generated
Recall performed		
Performed actions	10.82 (2.97)	12.45 (3.66)
Observed actions	9.82 (4.37)	10.91 (4.52)
Novel actions	24.00 (14.22)	
Recall observed		
Performed actions	10.23 (3.05)	13.77 (3.34)
Observed actions	9.68 (3.43)	12.77 (3.68)
Novel actions	20.50 (9.40)	

Note. Mean with Standard Deviations in brackets.

$12.88, p < .001, \eta^2_{\text{p}} = .308$. Participants were more likely to retrieve self-generated than partner-generated actions, the generation effect. The Task by Generation interaction was marginally significant, $F(1,42) = 3.26, MSe = 12.88, p = .078, \eta^2_{\text{p}} = .072$, with the effect of Generation larger in the Recall observed task, $F(1,21) = 16.03, MSe = 16.03, p = .006, \eta^2_{\text{p}} = .433$, than the Recall performed task, $F(1,21) = 16.03, MSe = 16.03, p = .063, \eta^2_{\text{p}} = .155$. There was a significant main effect of Source, $F(1,42) = 5.77, MSe = 7.98, p = .021, \eta^2_{\text{p}} = .121$. Overall performed actions were retrieved more frequently than observed actions, the enactment effect. None of the remaining interactions were significant, $F < 1$.

I looked separately at novel actions generated at retrieval. There was no significant evidence that novel actions were generated more frequently at retrieval during the recall of observed than of performed actions, $F < 1$

Monitoring Finally, I looked at the rates with which participants reported performed, observed and novel actions as task-compliant. The means are shown in Table 4.11. I first analysed actions generated at retrieval with a 2 Task (Recall performed, Recall observed) x 2 Generation (Partner-generated, Self-generated) x 2 Source (Performed, Observed) mixed ANOVA with repeated measured on the second and third factor. There were no significant main effects of Task, $F(1,42) = 1.69, MSe = 0.04, p = .20, \eta^2_{\text{p}} = .039$ and Source, $F < 1$. Unsurprisingly, there was a significant Task by Source interaction, $F(1,42) = 75.25, MSe = 0.07, p < .001, \eta^2_{\text{p}} = .642$. Participants in the Recall performed task reported performed actions more frequently than observed actions, $F(1,21) = 42.07, MSe = 0.07, p < .001, \eta^2_{\text{p}} = .667$, while the reverse was true in the Recall observed task, $F(1,21) = 33.93$, $MSe = 0.08$, $p < .001$, $\eta^2_{\text{p}} = .618$. There was a marginally significant main effect of Generation, $F(1,21) = 3.03, MSe = 0.02, p = .089, \eta^2_{\text{p}} = .067$, with an overall trend to report a

Table 4.11. Rates of actions generated at retrieval that were reported as task-compliant in Experiment 7

	Generation	
Source	Partner-generated	Self-generated
Recall performed		
Performed actions	0.77 (0.15)	0.70 (0.23)
Observed actions	0.37 (0.19)	0.38 (0.19)
Novel actions	0.52 (0.19)	
Recall observed		
Performed actions	0.43 (0.21)	0.41 (0.19)
Observed actions	0.79 (0.18)	0.75 (0.21)
Novel actions	0.64 (0.19)	

Note. Mean with Standard Deviations in brackets.

higher rate of partner-generated than self-generated actions. The remaining interactions were not significant, $F < 1.21, p > .27$.

I separately looked at the rate of novel actions reported as task-compliant. There was a marginally significant effect, $F(1,42) = 3.91, MSe = 0.04, p = .055, \eta_p^2 = .085$, indicating a trend that participants in the Recall observed condition were more likely to report novel actions as task-compliant than participants in the Recall performed condition.

4.5.3 Discussion

Unlike the previous experiments, Experiment 7 showed solid enactment and generation effects. Irrespective of participants' retrieval task, participants had more performed than observed actions, and more self-generated than partner-generated actions available to them at retrieval. While the previous two experiments showed strong effects of generating the ideas for performed actions on source memory, there was no evidence for an effect on source memory in this experiment. Generating the ideas of actions offered no cue towards the source of the action in this experiment nor was it necessary to consider that aspect of an item to complete the task. Participants were asked to recall self-performed or observed actions. In Experiment 5 and Experiment 6, identifying that an action had been self-generated, uniquely identified the item as self-performed and provided a benefit for the task. Identifying an item had been self-generated provided no diagnostic cue in Experiment 7. Rather than heuristically identifying self-generated actions as self-performed, participants therefore likely dismissed the cue as irrelevant. This suggests that participants can moderate their source responding based on the presence or absence of useful features that allow distinguishing sources of items.

4.6 Interim Discussion: Model-free analysis

In this chapter I tested the effects generating the idea for an action had on distinguishing performed from observed actions. In the experiments in the previous chapters, participants always generated and performed actions. So, can participants still identify the source of actions when they are asked to perform actions on instruction? While there was some evidence for participants being able to make use of the additional information generating an action provided, participants were still able to distinguish performed and observed actions when they had not generated the performed actions. This suggests that the effects reported in previous experiments were not solely driven by participants contrasting generated with non-generated ideas.

Part of the motivation for this chapter was to test the effects generation would have on item and source memory for action. In all three experiments, the extended recall task used at test showed dissociations of item and source memory after generation. In the first two experiments, generating the ideas for performed actions did not increase item memory for those actions as would be expected from a generation effect (Slamecka & Graf, 1978). This suggests that while enactment can enhance memory, and generation can enhance memory, both types of processing do not combine to create even stronger memories (see also Nilsson & Cohen, 1988).

This follows from literature that has shown consistently that enactment cannot be easily enhanced by further processing. R. L. Cohen (1983) showed that when single items' importance was stressed to participants, participants only showed a small retrieval benefit for those particular items. Neither participants' judgement-of-encoding-impact (R. L. Cohen, 1983) nor their judgement-of-trace-strength of an item (R. L. Cohen, 1988) was predictive of subsequent recall of self-performed actions. R. L. Cohen (1981) asked participants to engage in shallow versus deep encoding of self-performed actions by having them rate noisiness versus frequency of the action occurring. There was no effect of this manipulation on memory recall of self-performed actions. Helstrup (1987) varied the complexity of actions (shallow encoding: sorting playing cards into 2 stacks, complex encoding: sorting playing cards into 4 stacks). There was no evidence for an effect of more complex encoding on memory recall, though the author points out that the meaning of the action was the same in both versions of the task. While enriching action phrases by additional detail like "Wave your hands ... like a conductor" did not enhance recall of self-performed actions (Nilsson et al., 1995), enriching the action sequence like “Break the match... and pile the parts in front of you” impaired (rather than enhanced) recall of self-performed actions (Nilsson & Cohen, 1988). Zimmer and Engelkamp (1999) reported that conceptual processing (judge whether an action is appropriate) benefited memory for action phrases over surface processing (count letters). Yet additionally enacting actions did not significantly increase the advantage. These findings seem to suggest that once an item was encoded by enactment, its encoding cannot be further enhanced by an additional processing manipulation.

However, I did observe a clear generation effect (in addition to an enactment effect) when enactment and generation were fully crossed within-participants and within-sets in Experiment 7. This finding is in contrast to the null-generation-effect reported by Nilsson and Cohen (1988) for actions. They asked participants to read actions, generate actions, read-and-perform or generate-and-perform actions in a between-subjects design. They found no evidence for generate-and-perform actions being any more likely to be retrieved in a free recall task than read-and-perform actions. When I test the same comparison in Experiment 7 (across both retrieval tasks and ignoring the effect of generation on observed actions), generation does significantly improve memory for performed actions, $t(43) = 4.04, p < .001, d_{av} = 0.80$. This does suggest that the experimental design (within-participants and within-sets versus between-subjects) does impact the emergence of generation effects here. The experiments cited above all attempted to manipulate the strength of the enactment effects in between-subject manipulations. It is possible that the strength of additional processing only results in clear effects when those items directly compete against less processed items at retrieval in line with the conclusions of Bertsch et al. (2007) and with distinctiveness accounts (Hunt, 2012).

The effect of generation on source memory performance across the experiments is complementary to the item memory results. While Experiment 5 and Experiment 6 showed a benefit for source memory performance when performed actions had been generated, there was no clear benefit when both performed and observed actions were generated in Experiment 7. Those findings are most consistent with an account that assumes that participants use source information when it is diagnostic of a source in a particular task (only performed actions were generated) but not when it is not (performed and observed actions were both generated). Leynes and Kakadia (2013) proposed in line with Mitchell and Johnson (2000) that performed and observed actions are characterized by specific qualitative features (the diagnostic features account). If items contain those features, they are assigned to those sources. In this case, the cognitive feature associated with the item by generation should have led participants to favour ‘performed’ responses for self-generated items. There was no evidence for that. When the additional source manipulation was not uniquely associated with only one source, participants likely did not take this source feature into account at all. This provides evidence against a diagnostic features source monitoring account, but does not prohibit the interpretation that participants evaluate the source features when they are useful to make the source decision (in line with a more general source monitoring account).

One alternate account that is used to explain source memory performance is the relative strength account. According to this account, source decisions are made based on the overall memory strength of items, with stronger items assigned to the stronger source and weaker items assigned to the weaker source. Since I observed increases in source memory performance in the absence of item memory performance, this account struggles to explain the data. That said, there is a specific pattern in the data in Experiment 6 that may provide some support for this account. Participants in the Recall performed task showed a trend towards retrieving more generated than non-generated performed actions (i.e., retrieved stronger rather than weaker items), while participants in the Recall observed task showed the opposite trend towards retrieving the weaker performed over the stronger performed items. There are two possible reasons for this pattern of data. It is possible that participants chose not to report all items that came to mind as they performed the task and withheld the stronger memories of self-generated performed items because they knew they were not the target responses in their task. Alternatively, participants’ search strategy that results in a search set (that ideally would be reflected by the items retrieved at test prior to monitoring judgements) was focused on the weaker items since participants’ main task was to recall observed actions. This suggests that this interaction would be fully theoretically compliant with a relative strength account with a strong moderating influence of retrieval task orientation.

A final possibility is that the extended recall task may mask that participants’ item and source decisions are based on the same underlying memorial distributions by calculation of two separate indices and a recall task with two different retrieval task orientations. The two indices are not fully independent (and are calculated on different scales), so it is possible that the dissociation is an artefact of the analysis. I therefore also fitted formal models to the data in all three experiments.

4.7 Model-based analysis

Of the source memory accounts I described in the introduction, both the source monitoring account (in a less specific instantiation than the diagnostic features interpretation presented by Leynes and Kakadia (2013)) and the relative strength account are viable candidates to account for the pattern of data described across the three experiments. There was little evidence for the mistaken fluency account that would argue that participants are largely unable to orient towards a retrieval task and simply report the most fluent items every time.

Experiment 5 and Experiment 6 suggest a dissociation of item and source memory, with generation not significantly increasing the availability of performed actions at retrieval, but leading participants to make more correct source judgements. This potential dissociation maps onto a two-dimensional signal detection model of source memory typically described for source recognition paradigms (DeCarlo, 2003; Hautus et al., 2008; Johnson et al., 1993; Slotnick & Dodson, 2005). As described in Chapter 3, this model assumes separate item and source memory dimensions. When items exceed criteria on those dimensions, items are classified as old, and Source A respectively. At first glance, the data observed in the experiments above seems to favour that model. This account is closest to the diagnostic source features account (Leynes & Kakadia, 2013; Mitchell & Johnson, 2000) and more broadly typical source monitoring accounts. These accounts assume that item and source memory are dissociable.

The relative strength account discussed above maps onto a one-dimensional model of memory strength in recall tasks based on the one-dimensional signal detection model (Hollins et al., 2016; Marsh & Bower, 1993; Marsh & Landau, 1995). According to this model, responding is based on overall memory strength alone, with manipulations affecting the strength of distributions and/or placement of criteria for responding. At first glance, the observed data do not seem to provide evidence for the latter model since the data suggest a dissociation of item (availability) and source memory (monitoring). However, it is possible that participants' monitoring performance is based on monitoring the strength of sources (and items), rather than monitoring and comparing source features. In that case, stronger items would be more easily identified as performed and less likely to be falsely reported as observed. Under a one-dimensional model, changes in strength of distributions in combination with changes in decision criteria may be able to accommodate the data.

Signal detection models are typically described for recognition data. A recognition task results in classification of all encoded items by all participants. Items are explicitly classified as belonging to Source A or B or classified as New. Participants in a recall task only classify the items they recall, with participants showing differing retrieval rates. This means not all encoded items are classified at test in recall tasks. In a standard source-cued recall task, items are explicitly classified as Source A or not reported (or Source B and not reported). In an extended recall task, items are classified as Source A, Not Source A or not reported (or Source B and Not Source B). Neither standard source-cued recall or extended recall tasks provide sufficient data to fit complete signal detection models (though see Marsh & Bower, 1993, for a theoretical framework), but it is possible to estimate the best fitting parameters for the model given the information that is provided by the task using the least squares method.

I fitted both the one-dimensional model (representing the relative strength account) and the two-dimensional model (representing the source monitoring account) to the data for all three experiments. The underlying equations for the distributions and boundaries in the one-dimensional signal detection model and two-dimensional signal detection model are equivalent to those defined in Chapter 3. Figure 4.1 illustrates how data from selective recall and extended selective recall experiments can account for the decision space in the one-dimensional signal detection model (the data similarly accounts for the decision space in the two-dimensional signal detection model). The logic behind the analysis is described in full in Appendix C.

As Figure 4.1 suggests, I fitted both the final report recall data (that is, only items participants reported as task-compliant) and the extended recall data (that is, items participants reported and withheld from report). Given the experimental design, fitting the final report data meant combining the aggregate data of participants in the Recall performed and Recall observed task for a full estimation of the model. For the extended recall data, all participants provided all necessary information. However, even in this extended recall, participants are orienting towards a retrieval task. I estimated parameters across the aggregate extended recall data of participants from both retrieval tasks, assuming that the orientation towards observed actions in the Recall observed task and towards performed

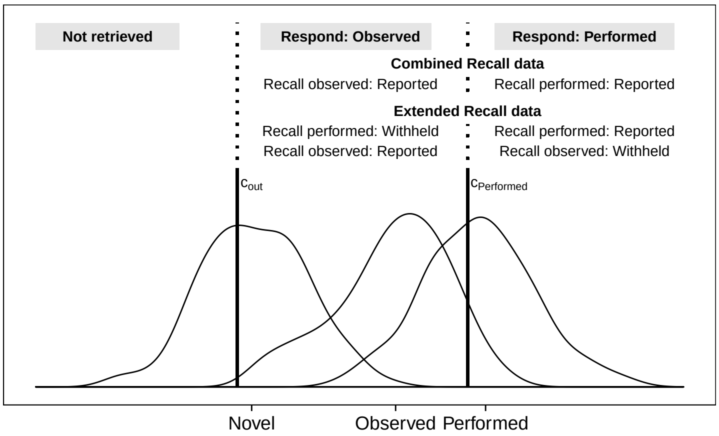

Figure 4.1. Illustration of the single-dimension model of source memory in a recall task for novel items and two sources of old items (performed actions and observed actions), with decision criteria are c_{out} and$c_{performed}$. At the top, illustration of the observed data the model is fitted to when fitting only the final report data or the full extended recall data

actions in the Recall performed task would even out and provide a more balanced fitting of the underlying memorial distributions.

I assumed equal-variance in both models. The parameters I am looking to fit in the one-dimensional signal detection model are the mean of the performed-actions distribution (D_P), the mean of the observed-actions distribution (D_O) and the placement of c_{old} and $c_{Performed}$. With the distribution of novel items set to 0, as per convention, the strength of the distributions of performed and observed actions will be calculated relative to that baseline. I adapted the two-dimensional signal detection model to recall data in the same way as the one-dimensional signal detection model. Since memory distributions are assumed to be bivariate not univariate in this model, the number of fitted parameters is higher. I defined an equal-variance case with independent dimensions. I will estimate parameters for the mean strength of performed and observed actions on the item dimension (D_P and D_O) and on the source dimension (d_P and d_O), as well as the decision criterion on the item dimension (c_{old}) and the decision criterion on the source dimension ($c_{Performed}$).

Models were fit using the *optim* function in R (using the conjugate gradient search method) by minimizing the residual sums of squares, separately for the data in Experiment 5 and Experiment 6. I discuss Experiment 7 separately below. Figures 4.2 and 4.3 show the observed data and the values predicted by the models (note, the extended recall data is plotted as number of actions reported and withheld, rather than number of actions produced at retrieval (availability) and proportion of actions reported as task-compliant (monitoring)). The best-fitting parameter estimates are in Tables 4.12 and 4.13.

Table 4.12. Parameter estimates of for models fitted to data in Experiment 5

Model	D_P	D_O	c_{out}	d_P	d_O	$c_{Performed}$
Combined recall data[1]						
1D						
Exp-pro	0.79	-0.72	-3.29			0.38
Par-gen	3.92	0.18	2.75			0.70
2D						
Exp-pro	1.95	1.49	-1.28	0.78	-0.68	0.38
Par-gen				1.03	-0.52	0.61
Extended recall data[2]						
1D-ext						
Exp-pro	0.84	-0.72	-3.25			0.41
Par-gen	4.18	0.20	2.91			0.75
2D						
Exp-pro	1.95	1.49	-1.28	0.83	-0.67	0.40
Par-gen				1.11	-0.59	0.56

[1] Participants' reported responses in both Recall performed and Recall observed task, combined across tasks

[2] Participants' reported and withheld responses, fitted simultaneously for both tasks

Exp-pro - Sets with experimenter-provided actions; Par-gen - Sets with participants-generated actions

I will first examine the figures and parameter estimates to see how well the models can account for the effects of generating actions reported in the model-free analysis. Both one-dimensional and two-dimensional signal detection models account well for the data, with the majority of predicted values falling within the 95% confidence intervals of the data. The two-dimensional model accounts for the various patterns of data and effects of the Generation manipulation on the basis of changes in average source strength of performed and observed actions. This provides some support for item and memory strength dissociating onto different dimensions in line with the source monitoring framework (Johnson et al., 1993). Interestingly, the one-dimensional model can account for that critical source-memory effect of self-generating performed actions (observed partially in Experiment 5 and fully in Experiment 6) as well, and does so on the basis of changes in overall memory strength and changes in criterion. This suggests that an account that assumes responding occurs on the basis of a single memory dimension can account for that particular pattern of responding, despite the seeming dissociation in the observed data.

I also fitted both models to the data in Experiment 7. Novel actions recalled at test could not be assigned to either Generation condition in this experiment, because Generation was manipulated within-set. To be able to fit the model, I therefore created a dummy-datapoint for novel actions in both Generation conditions by dividing the novel actions recalled at test by 2 for each participant and assigning them equally to both conditions.

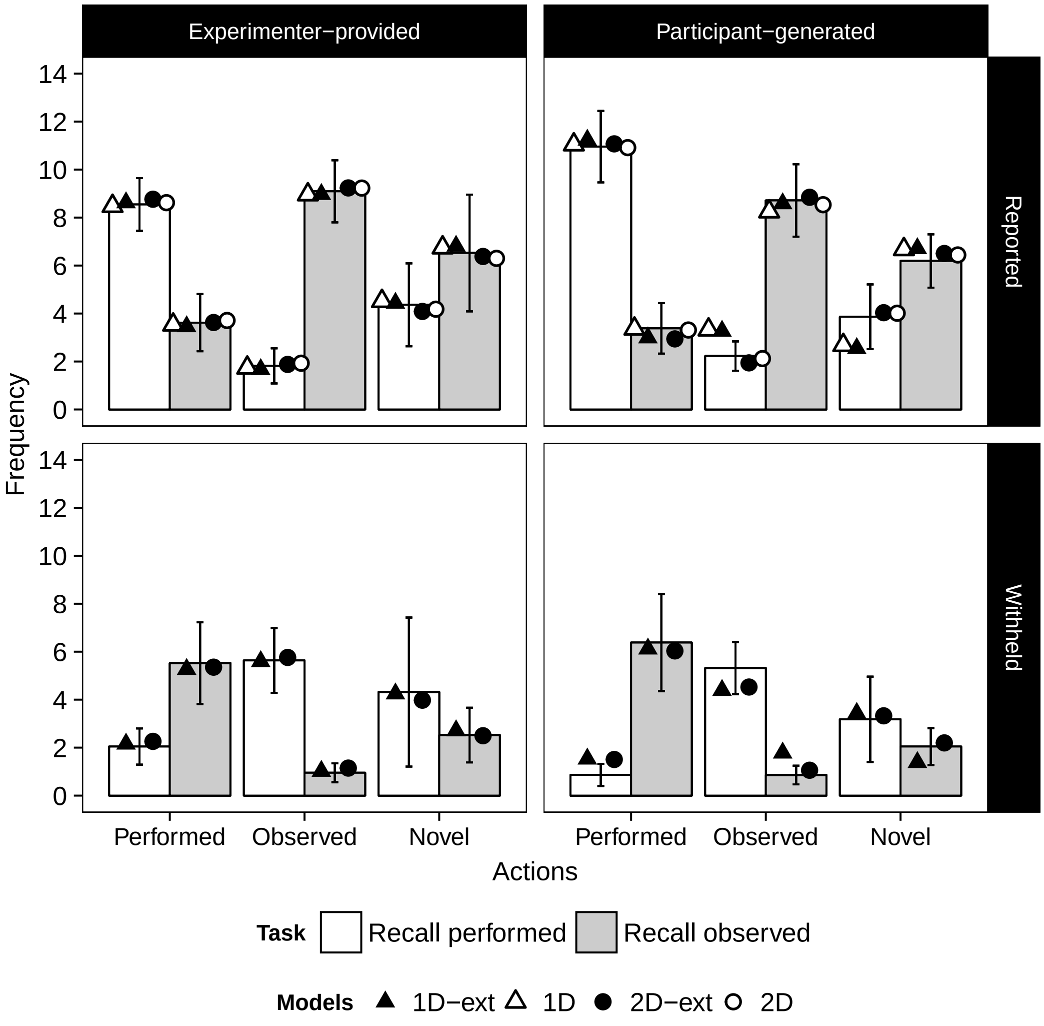

Figure 4.2. Observed and predicted frequencies during recall of actions in Experiment 5. The bars show observed frequencies. The errors bars are 95% confidence intervals. The fitted models are: 1D - one-dimensional signal detection model fitted to reported data, 2D - two-dimensional signal detection model fitted to reported data, 1D-ext - one-dimensional signal detection model fitted to reported and withheld data, 2D-ext - two-dimensional signal detection model fitted to reported and withheld data.

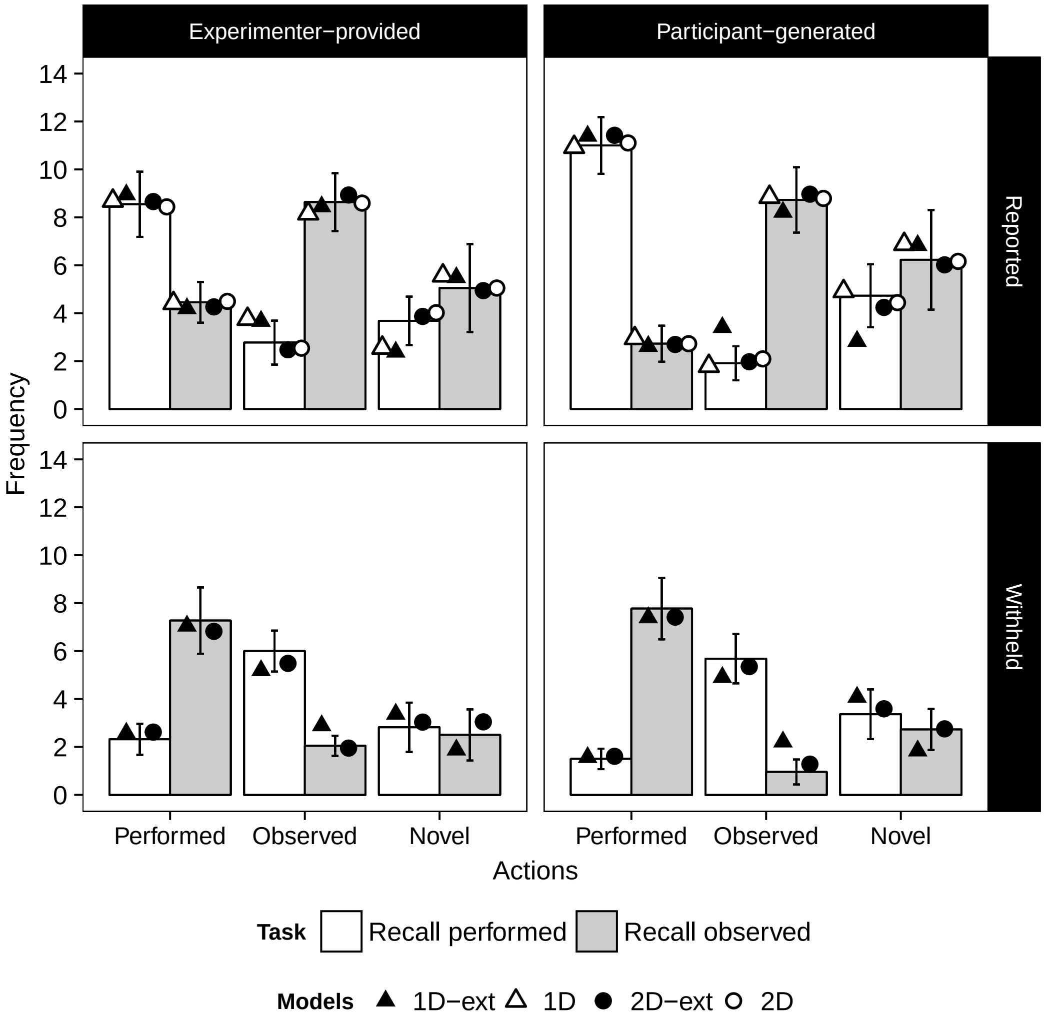

Figure 4.3. Observed and predicted frequencies during recall of actions in Experiment 6. The bars show observed frequencies. The errors bars are 95% confidence intervals. The fitted models are: 1D - one-dimensional signal detection model fitted to reported data, 2D - two-dimensional signal detection model fitted to reported data, 1D-ext - one-dimensional signal detection model fitted to reported and withheld data, 2D-ext - two-dimensional signal detection model fitted to reported and withheld data.

Table 4.13. Parameter estimates of for models fitted to data in Experiment 6

Model	D_P	D_O	c_{out}	d_P	d_O	$c_{Performed}$
Combined recall data[1]						
1D						
Exp-pro	3.85	0.37	3.12			0.54
Par-gen	1.02	-0.72	-3.09			0.47
2D						
Exp-pro	2.03	-0.39	0.05	0.37	-0.49	-1.63
Par-gen				0.90	-0.67	-0.75
Extended recall data[2]						
1D-ext						
Exp-pro	4.63	0.39	3.80			0.63
Par-gen	4.83	0.15	3.52			0.67
2D						
Exp-pro	2.02	-0.34	0.12	0.42	-0.53	-1.56
Par-gen				0.94	-0.71	-0.71

[1] Participants' reported responses in both Recall performed and Recall observed task, combined across tasks

[2] Participants' reported and withheld responses, fitted simultaneously for both tasks

Exp-pro - Sets with experimenter-provided actions; Par-gen - Sets with participants-generated actions

Beyond that, the fitting procedure and models fitted were identical. Observed and fitted data are shown in Figure 4.4, based on the parameter estimates in Table 4.14. Again, both models account well for the critical patterns of data.

Simply examining the figures for how they account for the effects of generation suggests that both models do quite well. This seems to suggest that both assuming responding occurs on the basis of overall memory strength (relative strength account) or on the basis of separable item and source memory dimensions (source monitoring account) could account for the effects of generation observed in the experiments. Closer examination of the figures for how well the models do beyond capturing those critical effects reveals that the two-dimensional model does better in capturing the full pattern of responding than the one-dimensional model. Both models struggle, in particular, to account for the high number of novel actions reported by participants in the Recall performed task relative to the number of observed actions, with the misfit more consistent across conditions for the one-dimensional model

This discrepancy in model fit is reflected in the formal model fit statistics. Table 4.15 shows the goodness of fit by residual sums of squares (RSS) and root mean square error (RMSE). The lower values for the two-dimensional models reflect the overall better fit of the two-dimensional models. Despite the two-dimensional model being more complex than the one-dimensional model, the two-dimensional model is still preferred over the

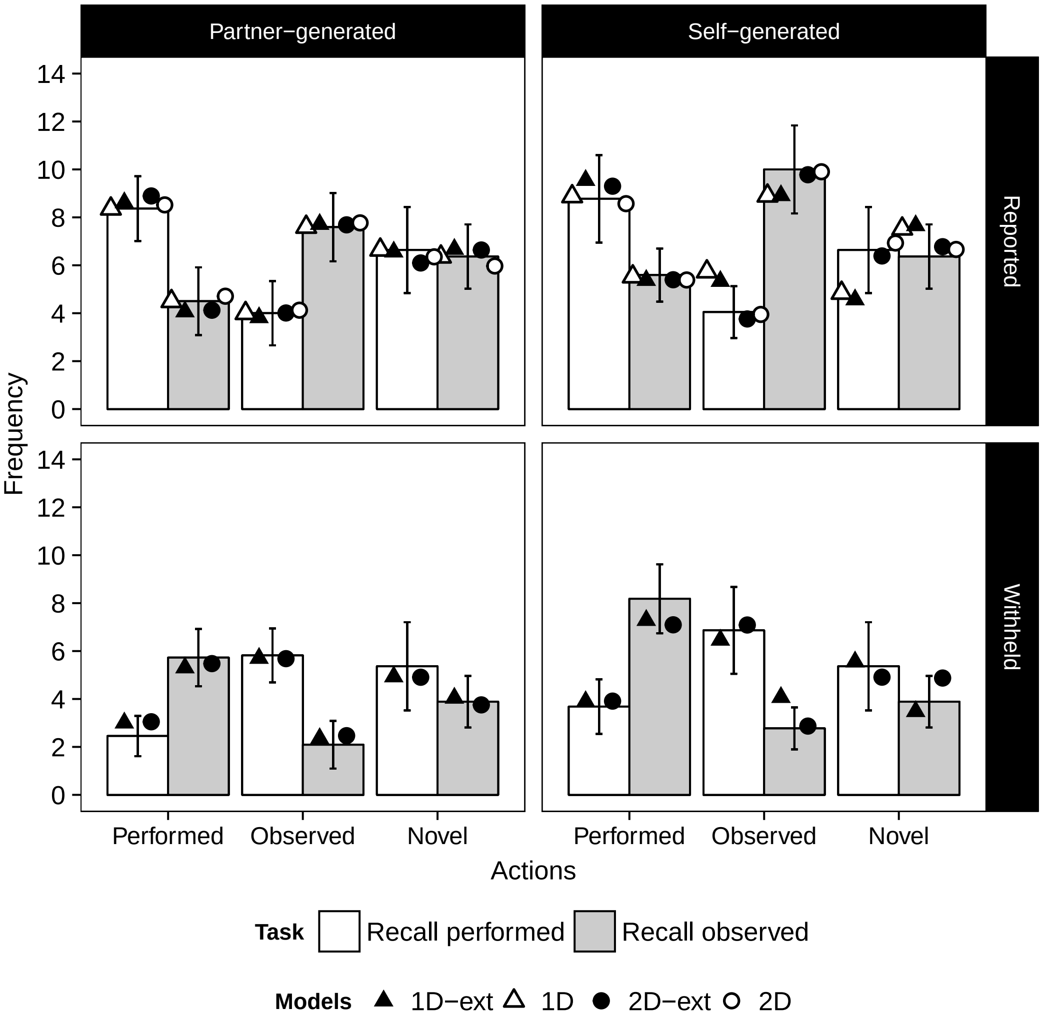

Figure 4.4. Observed and predicted frequencies during recall of actions in Experiment 7. The bars show observed frequencies. The errors bars are 95% confidence intervals. The fitted models are: 1D - one-dimensional signal detection model fitted to reported data, 2D - two-dimensional signal detection model fitted to reported data, 1D-ext - one-dimensional signal detection model fitted to reported and withheld data, 2D-ext - two-dimensional signal detection model fitted to reported and withheld data.

Table 4.14. Parameter estimates of for models fitted to data in Experiment 7

Model	D_P	D_O	c_{out}	d_P	d_O	$c_{Performed}$
Combined recall data[1]						
1D						
Exp-pro	0.44	-0.61	-0.97			-0.25
Par-gen	3.47	0.26	3.21			-0.12
2D						
Exp-pro	1.82	0.70	-1.96	0.33	-0.43	0.18
Par-gen				0.27	-0.59	-0.02
Extended recall data[2]						
1D-ext						
Exp-pro	0.59	-0.66	-0.80			-0.31
Par-gen	3.09	0.20	2.66			0.06
2D						
Exp-pro	1.87	-0.55	-0.62	0.42	-0.21	-1.15
Par-gen				0.31	-0.52	-1.02

[1] Participants' reported responses in both Recall performed and Recall observed task, combined across tasks

[2] Participants' reported and withheld responses, fitted simultaneously for both tasks

Exp-pro - Sets with experimenter-provided actions; Par-gen - Sets with participants-generated actions

one-dimensional model overall (by both AIC and BIC). It provides an overall better fit for the data, suggesting that source responding in a recall task is better captured by the assumption of two memory dimensions (item and source memory dimension).

That said, the source of the misfit of the models is identical for one-dimensional and two-dimensional models and identical for participant-generated and experimenter-provided conditions. All models struggle to account for the high number of novel actions reported by participants relative to the number of observed actions in the Recall performed task. Examining the parameter estimates shows why the misfits occur.

When the models follow the standard order of relative strength (performed actions have higher average strength than observed actions which in turn have higher average strength than novel actions), the models fail to account for novel actions being that prevalent in report. If observed actions are characterised by higher average strength, they should be reported more frequently. The model predicts that, and does not provide a good fit to the data. Since the models were not restricted to that order, for some of the conditions the models settled on a different order of average strength (performed actions have higher average strength than novel actions which have higher average strength than observed actions). In those cases the models actually provide a good fit to the pattern of data because they can account for the high prevalence of novel actions. I will discuss below what this means and why it may make sense to assume this less conventional order of

Table 4.15. Model fits for models fitted to all experiments in Generation

Model	Parameters	RSS	RMSE	AIC	BIC
Experiment 5					
Combined recall data[1]					
1D	8	3.18	0.51	29.88	3.94
2D	9	0.26	0.15	1.96	-23.50
Extended recall data[2]					
1D-ext	8	6.34	0.51	60.32	-6.53
2D-ext	9	2.09	0.30	35.68	-29.99
Experiment 6					
Combined recall data[1]					
1D	8	3.31	0.53	30.37	4.43
2D	9	0.32	0.16	4.17	-21.28
Extended recall data[2]					
1D-ext	8	15.40	0.80	81.62	14.77
2D-ext	9	2.11	0.30	35.94	-29.73
Experiment 7					
Combined recall data[1]					
1D	8	8.64	0.85	41.88	15.94
2D	9	0.63	0.22	12.40	-13.05
Extended recall data[2]					
1D-ext	8	13.53	0.75	78.52	11.67
2D-ext	9	4.75	0.44	55.39	-10.28

[1] Participants' reported responses in both Recall performed and Recall observed task, combined across tasks

[2] Participants' reported and withheld responses, fitted simultaneously for both tasks

1D - one-dimensional signal detection model
2D - two-dimensional signal detection model
RSS - Residual sums of squares, RMSE - Root mean square error
AIC - Akaike Information Criterion
BIC - Bayes Information Criterion

memory strength in recall tasks.

While the two-dimensional models appears to settle more frequently into this second, less conventional order (and has more flexibility with an additional dimension), the one-dimensional model does not. This causes the misfit, and causes the two-dimensional model to be preferred overall.

4.8 Discussion

In this chapter, I investigated different theoretical accounts of source memory performance by varying one source feature (cognitive processes) during encoding. The model-free analysis suggested that generating the ideas for actions improves participants' ability to distinguish performed and observed actions only if generation was manipulated for performed actions but not for both actions. This means if given additional source-identifying information for one source, participants will be able to make use of that information to aid in their decision about the source of an item. Interestingly, this effect on source memory appears to happen in the absence of an overall increase in memory strength by the additional information. This seems to provide preliminary support for a diagnostic features account suggesting that evaluation of source features but not the increase in overall memory strength drives source decisions. Further, participants were clearly able to orient their responding to the retrieval task, suggesting that a mistaken fluency account of source memory could not account for the results either.

The dissociation of item and source memory reversed when participants generated both self-performed and observed actions, with generation clearly benefiting availability of actions at test but not benefiting or indeed impairing source memory judgements. This does not provide any evidence for source memory judgements observed in Experiment 5 and Experiment 6 being the result of the cognitive trace being specifically linked to performed actions providing some evidence against the diagnostic features account that argues that the cognitive trace is specifically associated with internally-generated events. Instead generation here, when generated items are directly compared to non-generated items, seems to improve overall memory strength with no relative benefit for either source and hence no impact on source memory. This favours a relative strength account. Arguably, it is possible to account for the data using a more general source monitoring account that posits that participants judge the strength of source evidence for either source (without the strength of source evidence being related to overall memory strength or the specific features providing source evidence being relevant). The strength of source evidence would be equal for both performed and observed actions, hence leading to a null effect. Given there is a clear benefit of generation to memory strength, the relative strength account could explain the results equally well and provide a more parsimonious explanation.

Both the relative-strength and source monitoring account can be mathematically defined as models based in signal detection theory. I fitted both one-dimensional and two-dimensional models to the data to represent both theoretical accounts in all three experiments. In all three experiments, that two-dimensional model outperformed the one-dimensional model on fit and was preferred even when taking the higher number of parameters into account. This provides evidence in favour of the more complex source monitoring account of source memory that argues for source judgements being based on more than just overall memory strength. However, closer examination of the model fits revealed that benefit in fit were not necessarily based in the better fit of the Generation conditions.

The origin of the misfit is in the basic definition of the item memory dimension in both one- and two-dimensional models. Here novel actions are assumed to provide a baseline, with the memory strength of encoded items (performed and observed actions) measured relative to this baseline. Novel actions set at 0 form an arbitrary point to compare distributions against. Relative to novel actions at 0, both observing and performing actions result in representations of some memory strength relative to 0, assumed to exceed 0. Typically the memory strength of performed actions is higher than that of observed actions. A direct prediction of this model because of the order of distribution on the item memory strength dimension is that when recalling performed actions, participants should be falsely reporting more observed than novel actions. In fact, the reverse is the case in all three experiments.

During the parameter estimation, the mean of the distribution of observed actions in the one-dimensional model frequently exceeded 0 and resulted in misfits of this specific pattern of the data. Since the parameters for the mean of the distributions were not restricted to exceed 0, the best parameter estimates for some conditions set the mean strength of the distribution of observed actions to a negative estimate. In those cases, the one-dimensional model does provide a good fit for the pattern of data. This better fitting solution of assuming the strength of novel and observed actions to be reversed was not consistently chosen and resulted in more frequent misfits of the one-dimensional model than the two-dimensional model. Additionally, in the two-dimensional model the frequency of observed actions is the result of parameter estimations on two independent dimensions, so the model is more flexible in fitting this pattern of data.

The same mismatch of prediction and data was discussed in Hollins et al. (2016) for verbal category exemplars, so is not specific to the action stimuli used in this experiment. Accommodating that pattern within the theoretical contents of the mathematical framework is difficult. It is possible that a different set of assumptions (modelling t rather than gaussian distributions or modelling unequal variance models) could account for the pattern of data without requiring a reversal of memory strength. Alternatively, this pattern of data requires an examination of what memory strength means in a recall task.

Signal detection models are typically formulated for recognition data. Here all, or a subset of items, are chosen by the experimenter and classified by the participant. Some of those items are not items the participant would recall in a recall test. When participants' classify items as old their item memory strength exceeds the criterion, when they classify them as new, their item memory strength does not exceed this criterion. Here, the memory strength dimension can be easily defined as representing item memory strength. In a recall task, this memory strength dimension serves as a recall strength dimension more than a pure item memory strength dimension. Items are classified as old when they are recalled and as new, by default, if they are not recalled at all. Arguably, it is not just item memory that drives whether an item is recalled, but also an ability to access the item on cue. In fact, Marsh and Bower (1993) referred to this dimension as the 'Associative strength' dimension in a recall task, though they also posited that both encoded sources exceed novel items on this dimension.

What does it mean, then, that the models provide better fit here when the average strength of novel items is closer than the strength of observed actions to that of performed actions? Recall that novel actions in this task are actions participants generate at retrieval that they believe to have been generated by themselves. Novel actions are created by participants at retrieval, and in this particular experiment are likely recombinations of actions and objects that are close to correct responses without being quite correct. That those responses come to mind and are judged to be more similar to self-performed actions than those originating from an external source is quite plausible but does suggest something beyond item memory processes contributing to these decisions.

This suggests that conceptualizing this dimension as an associative strength dimension could better account for the pattern of data than thinking of it as an item memory strength dimension. However, the question then is what exactly drives associative strength? In a one-dimensional model of source memory, one could easily assume that source memory features are subsumed into this strength dimension and drive source recall on the basis of the relative strength of distributions on this composite trace. Extending this model to a two-dimensional model is a little less plausible, since source features would now be relevant for judgements on both dimensions.

In sum, while there was no evidence for the mistaken fluency account of source memory, both the relative strength account and the source monitoring account are likely candidates to account for the pattern of data. The more complex account proved to be more consistent mathematically in accounting for patterns in the recall data but both accounts were able to explain the effects of the generation manipulation. Both accounts struggled to explain that observed actions could be withheld more easily than novel actions when reporting performed actions, with the two-dimensional model more consistently explaining this pattern with additional variation in source strength. This highlights the need for testable models of source recall that address the specifics of the recall task rather than simply adapt models from source recognition paradigms.

After examining the impact of the motor component and the cognitive component, I will turn to a third source feature in the final empirical chapter. While participants in Experiment 5 observed their partners perform actions from across the table (third-person perspective), participants in Experiment 6 observed them from a first-person perspective. Observing performance from a first-person perspective increases the visual similarity of performed and observed actions, and should increase source confusion. Indeed there were some trends in the data that suggested that visual perspective may have affected source memory performance. I will explore this directly in the next chapter.

Chapter 5

Perspective

5.1 Background

In Chapter 2 and 3, I investigated the role of the motor trace in the source confusion of action memories, while I tested the impact of the cognitive trace in the experiments in Chapter 4 . The framework proposed by Leynes and Kakadia (2013) that I discussed in the introduction includes visual perspective as decisive identifier of performed and observed actions. Their argument for a first-person versus third-person visual perspective as a qualitatively distinguishing feature of performed and observed actions respectively is based on the results reported by Hornstein and Mulligan (2004) (I will discuss these below). In this framework self-performed actions are, if anything, characterised by a first-person perspective visual information. This visual cue, seeing hands from a first-person versus a third-person perspective should provide a strong cue in making decisions about the source of our action memories. According to a view that source features are specifically diagnostic of sources, watching own actions being performed from a third-person perspective should lead to participants disowning those actions more frequently while observing someone else from a first-person perspective should lead to participants claiming more of those actions as self-performed.

There is some neuropsychological evidence that supports the distinction of first-person and third-person perspective in the perception of actions. Brain regions of the extrastriate area (a brain region identified to be responsible for visual perception of bodies) respond to both images of body parts and video clips of actions from a first-person and third-person perspective (Jackson, Meltzoff, & Decety, 2006; Saxe, Jamal, & Powell, 2006). Relative to the third-person perspective, the first-person perspective more strongly activates the somato-sensory cortex for photos of body-parts (Saxe et al., 2006) and more strongly activates the motor cortex for video clips of intransitive actions (Jackson et al., 2006). Behaviourally, both Bortoletto et al. (2013) and Watanabe et al. (2013) reported that observation of actions from a first-person perspective facilitated imitation (shorter latencies) of simple finger movements and grips relative to a third-person perspective. Similarly, Bach, Fenton-Adams, and Tipper (2014) showed that somatosensory activation (as measured by latency to react to tactile sensation) was higher when observing painful compared to non-painful actions, only when participants watched videos from a first-person perspective. While these studies report an online influence of perspective on perception and action execution, Hornstein and Mulligan (2004) reported results in a memory study that can be interpreted as evidence for visual perspective moderating memory (specifically, source memory) performance.

Hornstein and Mulligan (2004) asked participants to take turns performing actions themselves and observing the experimenter perform actions in chunks of five actions each. Participants performed their own actions with their eyes closed, their eyes open (first-person perspective) or watched themselves perform the actions in a mirror (third-person perspective) in a between-subjects manipulation. Participants always observed the experimenter from a third-person perspective. After a two-week delay participants were asked to perform a 3AFC (performed, observed, novel) source recognition test. Hornstein and Mulligan measured rates of false source identification contingent on correct identification of performed or observed actions as old. They reported a significant interaction of source by perspective. Closer inspection showed that false identification of the sources did not differ for the eyes-closed (.09 for performed and .11 for observed actions) or the first-person perspective condition (.13 for performed and .12 for observed actions). However, when participants observed themselves perform actions in the third-person perspective, they were more likely to misattribute performed actions as observed (.18) than observed actions as performed (.10). In other words, when participants perceived their performed actions to look like observed actions, they were more likely to misattribute them to have been observed.

Leynes and Kakadia (2013) interpreted this as evidence for the third-person visual perspective being diagnostic of observed actions, with the first-person perspective diagnostic of performed actions. Hornstein and Mulligan (2004) favoured an explanation based on the absolute quantity of visual information available to participants at encoding (rather than the quality of the visual information in the different conditions). Both explanations make similar predictions for manipulations of visual perspective but the one favoured by Hornstein and Mulligan is less specific as to the nature of the manipulation. They argued that visual information is generally associated with events perceived in the world (external to the self), and that observed actions are purely external events. Thus they interpreted their results to show that an increase in visual information for performed actions increases participants' tendency to misattribute actions as observed. With this view, it is not that a first-person or third-person perspective are inherently diagnostic of performed or observed actions. Rather, visual information is more indicative of external than internal events and a third-person perspective provides more visual information overall.

In Chapter 4, I presented two experiments that were virtually identical bar the visual perspective participants had of their partner's performance (third-person perspective in Experiment 5 and first-person perspective in Experiment 6). In both experiments, I predicted an increase in source memory accuracy for performed actions if participants generated the actions they performed themselves. In Experiment 5, performed and observed actions shared the verbal label, concept and overall context but differed in the visual perspective. In Experiment 6 features of performed and observed actions overlapped maximally, with the visual perspectives now identical for performed and observed actions (both first-person perspective). Even though this is a cross experiment comparison, it allows some tentative insights into the effects of visual perspective overlaps or differences.

The full predicted effect of generating actions on source memory performance was only observed in Experiment 6. In Experiment 5, the data trended in the same direction but did not reach significance. In line with an argument for visual perspective as a feature, in Experiment 5, participants could distinguish actions not just by the cognitive ('I generated these actions') trace but also by the visual perspective. Participants could now rely on two features to guide their source decisions. This may have limited the impact of the cognitive trace. When visual perspective was not indicative of source in Experiment 6, on the other hand, participants had to rely more strongly on the cognitive trace for source decisions, and I observed the clear, predicted benefit of generation on source memory for performed actions. This suggests, like Leynes and Kakadia (2013) proposed, that visual perspective can be used as a source-diagnostic feature to distinguish performed and observed actions.

There are, to my knowledge, only a further two studies that manipulated the visual perspective in source memory for actions experiments. Lindner and Echterhoff (2015) manipulated visual perspective of performance and imagination of actions using the imagination inflation paradigm and Lindner et al. (2010) manipulated visual perspective of observation using the observation inflation paradigm.

The basic structure of the observation inflation and imagination inflation paradigm is identical. In the initial encoding phase, participants perform some action phrases and read other action phrases. The critical manipulation occurs in the second encoding phase. Here participants are presented with action phrases again (some old, some new) that they are now asked to observe being performed (observation inflation paradigm) or imagine being performed (imagination inflation paradigm). Some of the actions in the second encoding phase participants only observe or imagine once, others they observe or imagine repeatedly. After a delay, participants complete a source recognition test about the first encoding phase. They are presented with all actions from the experiment plus distractor actions and are asked to decide for each action whether they initially performed it, initially read it or whether it was not presented in the initial encoding phases.

In inflation paradigms, typically only the rates of ‘performed’ responses are analysed (e.g. Goff & Roediger, 1998; Lindner & Echterhoff, 2015; Lindner et al., 2010; Lindner & Henkel, 2015; Lindner et al., 2016, 2012; Schain et al., 2012). The observation inflation effect refers to the increase in rates of false ‘performed’ responses to actions that were observed in the second encoding phase relative to rates of false ‘performed’ responses to actions that were not observed in the second encoding phase. In other words, the observation inflation effect describes how much observing someone else increases false ‘performed’ responses relative to a general tendency to respond ‘performed’. The imagination inflation effect is derived analogously.

Lindner et al. (2010) tested participants on the standard observation inflation paradigm. They crucially manipulated the visual perspective of the actions participants observed in the second encoding phase. Participants were shown videos that showed an actor (from the neck down) perform an action as if the actor was sitting across the table from them (third-person perspective and the standard perspective in observation inflation experiments) or that showed an actor perform an action from a first-person perspective as if participants were looking down at their own hands. Lindner et al. also manipulated whether participants had their eyes closed or their eyes open during performance of own actions in the initial encoding phase. This resulted in three between-subjects groups. Group 1 was the standard group who performed their own actions with their eyes open and observed actions from a third-person perspective. Participants in Group 2 performed actions with their eyes open and observed actions from a first-person perspectives. Participants in Group 3 performed actions with their eyes closed and observed actions from a third-person perspective. There was no significant difference in observation inflation effect between groups (a null-effect, no Bayesian analysis). The authors interpreted non-significant trends in the data towards a higher inflation effect in Group 1 and 3 as evidence for an interpersonal mirroring account (under which a third-person perspective would lead to increased motor simulation during observation than a first-person perspective). Though note that this account is based only on extrapolation of the possible interpersonal function (e.g. Blakemore & Decety, 2001) of mirror neuron system activation.

Beyond Lindner et al. (2010)'s interpretation of the results, their data also do not provide positive evidence for the importance of visual perspective suggested by Leynes and Kakadia (2013)'s framework. Under this account, matching observed actions to performed actions by a first-person perspective should have increased participants' rates of false 'performed' responses. There was no evidence for that, with the manipulation resulting in a null-effect. It is possible that the videos Lindner et al. showed were not successful in creating a sense of self-performance due to being disembodied arms on a computer screen, and that a stronger real-life manipulation may increase identification of participants with the person they observe perform actions.

Lindner and Echterhoff (2015) reported a series of three imagination inflation experiments. In the first experiment, participants performed some actions with their eyes open (first-person perspective) or read some actions in the initial encoding phase. In the second encoding phase, participants were asked to either imagine themselves perform actions or imagine another person perform actions. The instruction for the other-person imagination included a screenshot from the videos used in the observation inflation experiments and showed an actor from the neck down with the object on the table in front of them. The instructions for the self-imagination showed a screenshot of only the object on the table without the actor present. Participants were not instructed on how to imagine themselves perform the actions.

While imagining themselves perform the action increased false 'performed' responses, imagining the other person perform actions did not result in a significant inflation effect.

In Experiment 2, participants observed themselves perform actions via a webcam in the initial encoding phase. In the second encoding phase, they again imagined themselves or the actor from the screenshot perform actions. Both imagination manipulations led to significant imagination inflation effects, with no significant difference between manipulations. Comparing these effects across studies seems to suggest that the third-person perspective imagination only elicited false memories of self-performance when performance was also seen from a third-person perspective. Consistent with a source monitoring framework account, this would suggest that increased similarity of sources leads to increased source confusion. Importantly, this comparison is across experiments with the critical comparison to the first-person perspective imagination not significant.

In the third experiment, Lindner and Echterhoff (2015) asked participants to perform actions themselves or observe the experimenter perform actions in the initial encoding phase. In the second encoding phase, participants were again instructed to imagine themselves perform actions or imagine the experimenter perform actions. Both factors were fully crossed between subjects. When participants observed the experimenter perform actions, imagining the experimenter perform actions led to a greater imagination inflation effect than when participants imagined themselves perform actions. This is in line with predictions from a visual perspectives account. On the other hand, when participants performed actions themselves, both imagining the self and imagining the experimenter perform actions led to imagination inflation with no difference between conditions (though the non-significant trend was towards higher inflation with self-imagination than other-imagination).

What this suggests is that visual perspective can play a role in source memory performance but does not necessarily do so consistently or decisively. Hornstein and Mulligan (2004) reported that matching the perspective of performed actions to that typically of observed actions increased source confusion. Lindner et al. (2010) reported that matching the perspective of observed actions to that typically of performed actions had no significant effect. Lindner and Echterhoff (2015) reported that imagining performing actions oneself always increased false 'performed' responses regardless of the perspective with which participants saw themselves perform actions in the experiment. Other-imagination only led to false 'performed' responses when participants imagined a person in the room perform actions or had seen themselves perform actions in a third-person perspective. While imagining observing someone perform actions led to more 'observed' responses than when participants imagined themselves perform actions, the opposite trend of an increase of false ‘performed’ responses under self- versus other-imagination was not significant.

In addition to exploring whether visual perspective can function as a feature that is diagnostic of performed and observed actions, I will also explore the observation inflation paradigm more closely in this chapter. I discussed above and in the introduction that the analysis of the 3AFC (performed, read, new) source recognition data in the observation inflation paradigm typically focuses only rates of ‘performed’ responses (Lindner et al., 2010; Lindner & Henkel, 2015; Lindner et al., 2016, 2012; Schain et al., 2012). The observation inflation effect shows that a) observing someone perform actions in the second encoding phase increases false ‘performed’ responses relative to false ‘performed’ responses given to actions that were not observed and b) observing someone perform actions increases correct ‘performed’ responses.

This limited analysis of ‘performed’ responses only, prohibits investigation of what leads to the increase of these responses. It is not clear how the remainder of responses (‘read’ or ‘new’) distribute over the different sources of items and manipulations. Since rates of both correct and false ‘performed’ responses increase, the observation inflation effect may be the result of a simple shift in response bias after observation. While Goff and Roediger (1998) discussed this possibility in the original imagination inflation study, none of the observation inflation papers discussed the data in that light. If inflation effects are response bias effects, this may lead to different strands of experimental exploration of the effect such as specific instructional, base rate or test format manipulations. A number of manipulations based on predicted effects of the source monitoring or motor simulation literature (perceptual similarity, instructions with a focus to the qualitative features, visual richness) did not lead to significant effects in the studies reported by Lindner et al.. Possibly understanding the inflation effect as a bias effect with lead to re-examination of the theoretical mechanisms underlying it.

5.2 Aims and predictions

The aims of this chapter are two-fold. First, the experiment in this chapter will investigate the role of visual perspective during observation on false memories of self-performance. According to a source monitoring account, observing someone else from a first-person perspective increases the similarity of self-performed and observed actions and should lead to more source confusion (Johnson et al., 1993). Similarly, a motor simulation account would argue that mirror neuron activation is stronger during observation of actions from a first-person perspective relative to a third-person perspective (Jackson et al., 2006; Jeannerod & Anquetil, 2008; Watanabe et al., 2013). If mirroring is crucial for false claims of self-performance, this increase in motor activation should lead to more false claims of self-performance. Lindner et al. (2010) could not show an effect of visual perspective, but this may be due to asking participants to observe videos of action performance. In the experiment in this chapter, I therefore asked participants to observe action performance in a face-to-face interaction.

Second, this experiment aims to look at the mechanisms underlying false memories in the inflation paradigm. I will use the inflation paradigm to a) replicate the analysis typically performed on the data that focuses on rates of 'performed' responses and b) extend the analysis to a more standard analysis of source recognition data to better understand the mechanisms underlying the inflation effect. Despite participants completing a 3AFC source recognition task, typically only 'performed' responses are analysed. This limited analysis prohibits investigation of what leads to an increase in false 'performed' responses after observation. Consequently, in addition to the standard analysis of 'performed' responses, I will analyse the data using the source recognition analysis and model-based analysis from Experiment 4 to investigate if response bias can account for the pattern of data. The basic paradigm of Experiment 8 is therefore based on Experiment 1 in Lindner et al. (2016), with the manipulation of visual perspective replacing the manipulation of visual richness in their experiment.

5.3 Experiment 8

5.3.1 Method

Participants

46 members of the public participated for payment of £8. 9 participants' data were excluded from the experiment (4 participants did not attend all experimental sessions, 3 participants did not follow instructions at encoding or retrieval and 2 participants' data was not recorded due to technical fault). The analysis is therefore based on 38 participants' data.

Materials

The stimuli comprised 60 simple action descriptions involving one or two objects and an action verb (e.g. ‘Tear the paper’). The actions were largely taken from Lindner et al. (2016), with 7 actions replaced by motorically similar verb-object action descriptions. The complete list of actions is listed in the Appendix D.1. The 60 actions were split into 12 blocks of five actions each to be assigned to the manipulation conditions in the experiment. Assignment of blocks was counterbalanced across experimental conditions and participants.

Procedure

The experiment consisted of two sessions. The first session comprised two phases. In Phase 1, participants were shown 30 action statements (that is 6 blocks of the total 12 blocks of actions). Participants performed half the actions, and only read the description of the action, without performing it, for the other half of the actions. On each trial, participants were first presented with the name of an object on a computer screen (e.g. sponge). The experimenter then placed the object in front of participants. After participants pressed the space bar, a ‘Perform’ or ‘Read’ instruction appeared on the screen. After pressing the space bar again, participants were presented with the action statement (e.g. ‘Squeeze the sponge’). If the instruction had been to perform the action, participants now performed the action with the object in front of them once. If the instruction had been to read the action, participants merely read the action statement. Pressing the space bar again started the next trial. Order of ‘Perform’ and ‘Read’ trials was randomised across the experiment. Assignment of ‘Perform’ and ‘Read’ instructions and (blocks of) actions were counterbalanced across participants.

In Phase 2 of the encoding session, participants observed the experimenter perform actions. The actions the experimenter performed were actions participants had performed (10 actions), participants had only read (10 actions) and 10 novel actions. For each trial in this phase of encoding, participants first saw the name of the object, followed by the action phrase on a screen. They then observed the experimenter perform the action once, before initiating the next trial. Half of the actions were presented once (15 actions, made up to equal parts of performed, observed and novel actions), half were presented five times (15 actions). The order of actions was randomized, with the constraint that an action was not presented twice in a row. Participants were asked to note for each action whether it was performed with one or two hands, to ensure they paid attention throughout this encoding phase. To manipulate perspective of observation, half the participants sat across the table from the experimenter (third-person perspective) while half sat next to the experimenter (first-person perspective).

Participants returned for the test phase 2 weeks later. They were presented with a 3AFC source recognition test referring to the first encoding phase only. 60 action phrases were presented to participants, 20 of which were novel at test. For each trial at test, participants were shown the action phrase on the screen, with three response options at the bottom of the screen. Participants were asked to indicate whether they had performed the action in the first encoding phase, had only read it in the first encoding phase, or whether it had not been presented in the first encoding phase. The instruction given to participants prior to the experiment stressed that the source recognition test referred to the first encoding phase only and that they were to ignore whether or not they had seen the action being performed by the experimenter. Participants were explicitly informed that the correct identification of an action that had not been presented to them in the first encoding phase but that the experimenter had performed in Phase 2 was ‘Not presented’ rather than ‘Performed’.

5.3.2 Model-free analysis

I analysed the results of the experiment using a model-free and model-based analysis as in Chapter 3. The rates with which participants responded ‘Performed’, ‘Read’ and ‘New’ to items by observation frequency and visual perspective is given in Table 5.1. I will first report and then discuss the results of the model-free analysis. In this model-free analysis I will begin by analysing the data as Lindner et al. (2016) have done (the observation inflation analysis) and focus on the rates of ‘Performed’ responses. Following that, I will analyse the data with the approach to source recognition experiments I presented in Experiment 4. Here item memory and source memory are analysed separately on the basis of data from the 3AFC (performed, read, novel) source recognition task.

Observation inflation

The first analysis follows the observation inflation paradigm analysis (Lindner & Echterhoff, 2015; Lindner et al., 2010, 2016). This analysis focuses exclusively on the effects of observation and the visual perspective manipulation on the mean proportions of participants’ ‘Performed’ responses for each item type. Correct ‘performed’ responses are

Table 5.1. Rates of responses in the 3AFC task by source and visual perspective in Experiment 8

	Third-person			First-person		
	Performed	Read	Novel	Performed	Read	Novel
Not observed						
‘Performed’	0.46	0.07	0.03	0.61	0.07	0.02
	(0.28)	(0.16)	(0.05)	(0.23)	(0.12)	(0.05)
‘Read’	0.27	0.40	0.09	0.20	0.45	0.13
	(0.24)	(0.21)	(0.11)	(0.13)	(0.23)	(0.15)
‘Novel’	0.26	0.53	0.88	0.19	0.48	0.85
	(0.27)	(0.24)	(0.15)	(0.16)	(0.27)	(0.17)
Observed once						
‘Performed’	0.80	0.16	0.10	0.77	0.13	0.04
	(0.26)	(0.17)	(0.19)	(0.22)	(0.16)	(0.09)
‘Read’	0.10	0.56	0.31	0.11	0.53	0.30
	(0.15)	(0.26)	(0.17)	(0.15)	(0.22)	(0.34)
‘Novel’	0.10	0.25	0.59	0.11	0.33	0.66
	(0.16)	(0.20)	(0.27)	(0.20)	(0.25)	(0.35)
Observed five times						
‘Performed’	0.78	0.21	0.24	0.84	0.25	0.12
	(0.18)	(0.23)	(0.23)	(0.15)	(0.21)	(0.15)
‘Read’	0.12	0.59	0.29	0.09	0.56	0.35
	(0.13)	(0.26)	(0.28)	(0.15)	(0.22)	(0.32)
‘Novel’	0.10	0.20	0.47	0.07	0.19	0.53
	(0.12)	(0.20)	(0.27)	(0.10)	(0.18)	(0.33)

Note. Average performance by item (Performed, Observed and Novel in the columns) and response type (‘Performed’,‘Observed’ and ‘Novel’). Proportions of responses for each item type sum to 1, in both groups and perspectice conditions. Data is based on 19 participants in each group.

'performed' responses to actions participants had performed in the first encoding phase, while false 'performed' responses are 'performed' responses to actions participants had 'read' in the first encoding phase or that had not been presented at all. Lindner et al. typically collapse 'performed' responses over read and novel responses (as has been the case in all their reported experiments) "because the same pattern emerged for the two kinds of nonperformed items (read and not presented) from Phase 1" (Lindner et al., 2010, p. 1293).

In this experiment, findings also did not differ for read and not presented actions, so I collapsed across those two types of data. I analysed the resulting data with a 2 Initial encoding (Performed, Not performed) x 3 Observation frequency (0, 1, 5) x 2 Perspective (Third, First) mixed ANOVA with repeated measures on the first and second factor. Unsurprisingly, there was a main effect of Initial encoding, $F(1,35) = 495.37, MSe = 0.04, p < .001, \eta_{\text{p}}^2 = .934$. Participants correctly identified a higher proportion of performed actions as 'performed' than they falsely identified not performed actions as 'performed'. There was a significant main effect of Observation frequency, $F(2,70) = 30.48, MSe = 0.03, p < .001, \eta_{\text{p}}^2 = .465$. This means that observing actions increased the proportion of correct and false 'performed' responses. The main effects were moderated by a significant Initial encoding by Observation frequency interaction, $F(1,22) = 9.09, MSe = 0.21, p = .006, \eta_{\text{p}}^2 = .292$. The main effect of Perspective was not significant, $F(1,35) = 0.11, MSe = 0.06, p = .74, \eta_{\text{p}}^2 = .003$. The remaining interactions were not significant, all $F < 2.35, p > .13$.

I looked at the Initial encoding by Observation frequency interaction separately with multiple bonferroni-adjusted pairwise two-sided t-tests, collapsed across the perspective conditions. Observing the experimenter perform actions five times increased both false 'performed', $t(36) = 5.94, p < .001, d_{av} = .37$ and correct 'performed' responses, $t(36) = 5.44, p < .001, d_{av} = .32$, relative to baseline of not observing the experimenter perform actions. The same was true when the experimenter performed the action only once for both false 'performed' responses, $t(36) = 4.31, p < .001, d_{av} = .21$, and correct 'performed' responses, $t(36) = 4.70, p < .001, d_{av} = .25$. While false 'performed' responses significantly increased with additional repetition (comparing one versus five observations), $t(36) = 3.42, p = .009, d_{av} = .15$, the increase was not statistically significant for correct 'performed' responses, $t(36) = 0.60, p = 1, d_{av} = .03$.

Figure 5.1 shows the inflation rates in the experiment, that is the difference of false 'performed' responses to actions that have been observed and actions that have not been observed during the experiment.

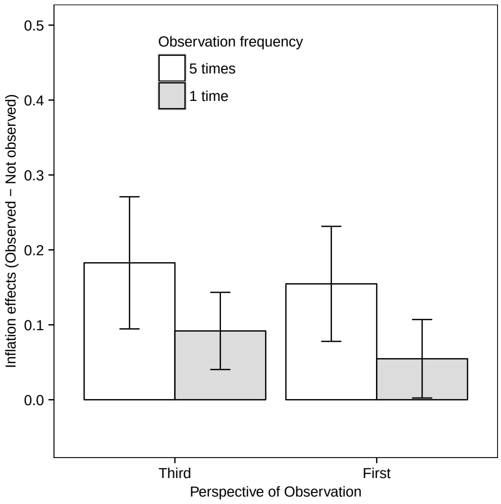

Figure 5.1. Observation inflation rates after single and repeated observation by visual perspective condition in Experiment 8. Error bars are 95% between-subjects confidence intervals.

In sum, observing previously encoded actions increases both false 'performed' responses and correct 'performed' responses at test. Critically, the perspective of observation has no significant effect on the observation inflation effect. This pattern of data was identical (bar a different between-subjects manipulation) to the pattern of data reported by Lindner et al. (2016).

3AFC analysis

Lindner et al. (2010) report only and focus exclusively on 'performed' responses resulting from their data. While they show consistently that 'performed' responses increase with observation following encoding, they do not explore the origin of this shift in responding. Increased proportions of 'performed' responses must be equivalent to decreased proportions of the other two response options. Critically, one of them would imply a shift in source attribution ('Read' responses) while the other implies a shift in item memory ('Novel' responses).

In addition to the observation inflation analysis above, I therefore also analysed the data using the same approach as for Experiment 4 in Chapter 3.

Item memory I analysed items correctly identified as old (that is, 'performed' or 'read') and novel items falsely identified as old. Since those measures can be subject to influence of response bias, I also looked at the data using the item discrimination parameter

Table 5.2. Item memory in Experiment 8

	Third			First		
Actions	0	1	5	0	1	5
Hits						
Performed	0.74	0.90	0.90	0.81	0.88	0.93
	(0.27)	(0.16)	(0.12)	(0.16)	(0.20)	(0.10)
Read	0.47	0.75	0.80	0.52	0.67	0.81
	(0.24)	(0.20)	(0.20)	(0.27)	(0.15)	(0.18)
False Alarms	0.12	0.41	0.53	0.15	0.33	0.47
	(0.15)	(0.27)	(0.27)	(0.17)	(0.34)	(0.33)
d'						
Performed	2.54	1.79	1.41	2.40	2.14	1.82
	(1.20)	(1.45)	(1.30)	(1.20)	(1.64)	(1.14)
Read	1.59	1.05	1.01	1.48	1.24	1.27
	(1.12)	(1.39)	(1.06)	(1.17)	(0.92)	(1.03)
Response bias c						
Performed	0.43	-0.64	-0.77	0.15	-0.43	-0.74
	(0.87)	(0.47)	(0.55)	(0.53)	(0.73)	(0.76)
Read	0.90	-0.28	-0.57	0.60	0.02	-0.46
	(0.79)	(0.55)	(0.75)	(0.70)	(1.02)	(0.87)

Note. Means with standard deviations in brackets.

d' and the response bias parameter c. The means for hit rates, false alarms, d' and c are in Table 5.2.

I first analysed item hit rates (that is, correct identification of old items as old) using a 2 Source (Performed, Read) x 3 Observation frequency (0, 1, 5) x 2 Perspective (Third, First) ANOVA with repeated measures on the first two factors. There was a significant main effect of Source, $F(1,35) = 59.11, MSe = 0.03, p < .001, \eta_{\text{p}}^2 = .628$. Participants' item memory hit rates were higher for performed than for read actions. There was a significant main effect of Observation frequency, $F(2,70) = 25.27, MSe = 0.04, p = .065, \eta_{\text{p}}^2 = .419$. The main effects were moderated by a significant Source by Observation frequency interaction, $F(2,70) = 3.57, MSe = 0.04, p = .033, \eta_{\text{p}}^2 = .093$. I looked separately at the effect of Observation frequency for performed and read actions. There was a significant main effect of Observation frequency for performed actions, $F(2,70) = 7.58, MSe = 0.03, p = .002, \eta_{\text{p}}^2 = .178$. Bonferroni-adjusted pairwise comparisons show that observing an action once $(p = .023, d_{av} = 0.61)$ increases correct identification of performed actions as old relative to baseline. Repeated observation, while leading to an increase relative to baseline $(p = .003, d_{av} = 0.85)$, does not lead to an additional increase of correct item identification relative to a single observation $(p = 1, d_{av} = 0.15)$. The effect of observation frequency was also significant for actions read at encoding, $F(2,70) = 18.92, MSe =$

$0.05, p < .001, \eta_p^2 = .351$. The pattern was identical to performed actions, in that both single ($p < .001, d_{av} = 0.94$) and repeated observation ($p < .001, d_{av} = 1.43$) increased correct identification of read actions as old, while repeated observation did not significantly boost identification further than single observation ($p = .20, d_{av} = .45$). The effect sizes show that read actions benefited from observation more than performed actions, likely due to performed actions already being frequently correctly identified in the baseline condition. The remaining main effects and interactions were not significant, all $F < 1$.

I next analysed the false alarm rates, that is false identification of novel actions as old (performed or read) with a 3 Observation frequency (0, 1, 5) x 2 Perspective (Third, First) mixed ANOVA with repeated measures on the first factor. There was a significant effect of Observation frequency, $F(2,70) = 53.52, MSe = 0.05, p < .001, \eta_p^2 = .414$. Bonferroni-adjusted pairwise two-sided comparisons showed a pattern identical to the one reported for hit rates. Single ($p < .001, d_{av} = 1.07$) and repeated observation ($p < .001, d_{av} = 1.64$) increased false alarms relative to baseline, but repeated observations did not additionally increase false alarms relative to single observations ($p = .21, d_{av} = .42$).

I analysed d' using a 2 Source (Performed, Read) x 3 Observation frequency (0, 1, 5) x 2 Perspective (Third, First) ANOVA with repeated measures on the first two factors. Participants showed higher discrimination of performed from novel than read from novel actions, $F(1,35) = 53.52, MSe = 0.55, p < .001, \eta_p^2 = .605$. There was a significant main effect of Observation frequency, $F(2,70) = 3.82, MSe = 0.96, p = .027, \eta_p^2 = .098$. Bonferroni-adjusted pairwise comparisons showed that observing performed, read and novel actions five times after initial encoding decreases item identification relative to not observing those actions ($p < .001$, $d_{av} = 0.55$). Similarly observing an action once after initial encoding decreased item identification (marginally significant at $p = .05, d_{av} = 0.36$). There was no evidence for observing the action five times significantly decreasing item identification compared to observing the action once ($p > .99, d_{av} = 0.14$). The Source by Observation frequency was not significant, $F(2,70) = 1.76, MSe = 0.56, p = .18, \eta_p^2 = .048$. All remaining main effects and interactions were not significant, $F < 1$.

I finally looked at the response bias c. There were significant main effects of Source, $F(1,35) = 53.52, MSe = 0.14, p < .001, \eta_p^2 = .605$, and Observation frequency, $F(2,70) = 37.95, MSe = 0.67, p < .001, \eta_p^2 = .520$. Participants were more liberal in responding to performed than to read actions, and participants became more liberal in their responding with increasing observation frequency (all bonferroni-adjusted pairwise comparisons $p < .012, d_{av} > 0.40$).

In summary, item detection (or item memory) decreases with observation. Possibly unsurprisingly, participants are better at discriminating items they encoded from entirely novel items, than discriminating items they encoded in the first encoding phase from items they encoded in the second encoding phase. At the same time, observing actions in the second encoding phase, makes participants more likely to respond ‘old’ (here: encoded in the first encoding phase) overall, i.e. encoding items by observation in the second encoding phase biases responding.

Source memory As in Chapter 3, I looked at participants’ source memory performance next. Participants’ source discrimination is calculated as correct source identification contingent on items having been correctly identified as ‘old’. The proportions of performed and read actions correctly identified by their source are reported in Table 5.3. I analysed these source hit rates with a 2 Source (Performed, Read) x 3 Observation frequency (0, 1, 5) x 2 Perspective (Third, First) mixed ANOVA with repeated measures on the first and second factor. Some participants classified all performed and read actions as novel at test (hence not providing any data for this analysis), so I analysed the data with a mixed model ANOVA using the *afex* package in R (Singmann et al., 2015). All post-hoc pairwise comparisons are bonferroni-adjusted and two-tailed. There was a significant Source by Observation frequency interaction, $F(2, 172.34) = 13.74, p < .001$. The remaining main effects and interactions were not significant, all $F < 1$. I looked at the effect of Observation frequency separately for performed and read actions. There was a significant effect of Observation frequency for correct source identification of performed actions, $F(2, 69.36) = 10.81, p < .001$. Participants’ source hit rates for performed actions increased with five action observations ($p < .001, d_{av} = 1.01$) and one action observation ($p = .002, d_{av} = 0.83$) relative to baseline. There was no evidence for a further increase in correct source identification from one to five action observations ($p = 1, d_{av} = 0.12$). For correct source identification of read actions, observation frequency showed the reverse trend, $F(2, 68.74) = 5.11, p = .009$. Observing actions five times ($p = .014, d_{av} = 0.71$) significantly decreased correct source identification relative to baseline. Neither the increase from baseline to one action observation ($p = .20, d_{av} = 0.42$) or one to five action observations ($p = .49, d_{av} = 0.32$) resulted in a significant decrease in source hit rates for read

Table 5.3. Source memory in Experiment 8

	Third			First		
Actions	0	1	5	0	1	5
Source hits						
Performed	0.63 (0.29)	0.87 (0.21)	0.87 (0.16)	0.73 (0.20)	0.85 (0.18)	0.90 (0.15)
Read	0.88 (0.23)	0.78 (0.21)	0.74 (0.16)	0.90 (0.20)	0.82 (0.18)	0.70 (0.24)
Source discrimination						
d'	1.34 (1.01)	2.25 (0.98)	2.13 (0.99)	1.56 (0.91)	2.08 (1.02)	2.15 (0.77)
Response bias c	0.25 (0.64)	-0.28 (0.66)	-0.23 (0.65)	0.03 (0.58)	-0.18 (0.53)	-0.44 (0.65)

Note. Means with standard deviations in brackets.

actions.

I additionally computed d' and c to combine responses to both sources into an index of source discrimination and a separate index of response bias. I analysed both measures using the same mixed ANOVA approach as above. There was a main effect of observation frequency on source discrimination of performed and read actions, $F(2,68.82) = 6.24, p = .003$. Bonferroni-adjusted pairwise comparisons showed that observing actions five times ($p = .007, d_{av} = 0.76$) or one time ($p = .009, d_{av} = 0.77$) increased source discrimination relative to baseline, while there was no evidence for a further change between one and five observations ($p = 1, d_{av} = 0.05$) There was no significant main effect of Perspective, nor was there a significant interaction, both $F < 1$.

At the same time, there was a significant main effect of observation frequency on response bias c, $F(2,68.74) = 5.82, p = .005$. Participants were more likely to respond ‘performed’ than ‘read’ when they observed encoded actions five times ($p = .011, d_{av} = 0.74$) or one time ($p = .017, d_{av} = 0.65$) relative to baseline. There was no evidence for a further increase in bias with increase from one to five observations ($p = 1, d_{av} = 0.11$). There was no significant main effect of Perspective, nor was there a significant interaction, both $F < 1$.

5.3.3 Discussion Model-free analysis

The aims of this experiment were two-fold: (1) I investigated if observing someone perform actions from a first-person perspective increases false ‘performed’ responses relative to observing someone perform actions from a third-person perspective, and (2) I analysed the source recognition data more fully, beyond the focus on rates of 'performed' responses to investigate the mechanisms underlying the results.

For (1), there was no evidence for an effect of visual perspective on responding in the source recognition test. While observation in the first-person perspective should have provided visual input during observation that is very similar to performed actions and thereby increase both mirroring (under a motor simulation account of observation inflation) and source misattributions (under a source monitoring account of observation inflation), there was no behavioural evidence for such an effect. This replicates the null-effect of visual perspective reported by Lindner et al. (2010). The change of action performance observed in a video to action performance observed in a face-to-face interaction does not appear to be crucially responsible for the null-effect of visual perspective in this paradigm.

In contrast, Hornstein and Mulligan (2004) reported an effect of visual perspective. Aside from a change in paradigm (and in action items), there is a crucial difference in the experimental design and results that may account for the difference in effects. While Lindner et al. (2010) and I manipulated the perspective of action observation, Hornstein and Mulligan manipulated the perspective of action performance (though note that Lindner and Echterhoff (2015) did not observe an effect with that manipulation). While Lindner et al. and I could not shown an increase in false 'performed' responses, Hornstein and Mulligan reported an increase in false 'observed' responses. In other words, it is possible that visual perspective influences the size of the error in one but not the other direction. There may be cues aside from the visual cue that are necessary for false identification of actions as performed (i.e., performed actions are defined by a richer set of features not easily emulated just by visual perspective), while the visual cue is sufficient to disown performed actions (i.e., the visual cue is relatively more important for the identification of observed actions). Alternatively, it is possible that any effect of visual perspective may not come to bear in an inflation paradigm, since participants do not explicitly make source decisions between performed and observed but performed and read actions.

For (2), the observation inflation analysis showed the same pattern of data as analyses reported by Lindner et al. (2016). This means I replicated the basic observation inflation effect as reported in the literature with face-to-face interactions, where observing someone else perform an action leads to an increase in subsequent 'performed' responses for both correct and false claims of performance. In addition to the standard observation inflation analysis, I also looked at the entirety of the data (beyond ‘performed’ responses) to trace the origin of the pattern of responding.

When looking at d' and response bias measures to separate memory and bias in responding, observing someone else perform an action overall decreases item memory (that is, discrimination of performed and novel actions), but increases source discrimination between performed and read actions. This seems paradoxical at first, but in addition to those memory effects, participants show a clear response bias to respond ‘old’ for performed actions and ‘performed’ during source discrimination after observation relative to baseline . This seems to suggest that observing someone else perform actions orients participants towards ‘performed’ responses above everything else.

One possibility is that read actions that were subsequently observed are now perceptually enriched. This new increased sensory overlap may drive the false responding (Thomas et al., 2003), in line with a source monitoring account that assumes that source features drive source memory performance (Johnson et al., 1993; Leynes & Kakadia, 2013). In contrast, Goff and Roediger (1998) argued for a memory strength based account. Under this explanation, observing someone else perform an action likely boosts memories for actions participants had performed themselves, and increases hit rates. At the same time, observation increases the familiarity with the action, and may lead to participants falsely attributing this action to the source that most resembles observed actions (and give a ‘performed’ response).

This latter explanation would suggest that observation inflation largely occurs because of a shift in response bias towards ‘performed’ responses. In that case, the data should be able to be accounted for by formal response bias models, i.e., models that assume that memory strength remains stable across the observation manipulation while guessing or decision criteria change.

5.3.4 Model-based analysis

The previous analysis showed a significant influence of response bias on performance in the memory task. Observing someone else perform actions appeared to increase ‘Performed’ responses relative to identifying actions as novel and relative to identifying actions as presented but not performed. The model-free analysis showed some changes in memory indices aside from the shifts in bias. In this model-based analysis, I tested if response bias models can account for the pattern of data entirely without necessitating changes in memory parameters. It is inarguable that observing someone else perform an action increases performed responses relative to the other two response options, but what are the mechanisms underlying this effect?

I chose to fit a reduced version of the data to reduce some of noise introduced by the experimental conditions. Given there was no evidence for an effect of visual perspective, I collapsed data across that manipulation and fitted the same model to all participants. Observation frequency was manipulated on three levels in the experiment. Some of the actions participants responded to at test they had not observed the experimenter perform at all, some they had observed the experimenter perform once, and some they had observed the experimenter perform 5 times. In all analyses, actions observed once resulted in similar, though (non-significantly) weaker effects relative to baseline as actions observed five times. For greater clarity, I only fitted the two extreme conditions, i.e. responses to actions participants did not observe and actions they observed repeatedly. The restrictions to the fitted data mean that the data is characterized by 18 data points (and 12 degrees of freedom).

As in Chapter 3, I fitted the 2-High-Threshold model (2HTM), the one-dimensional signal detection model (1D-SDT) and the two-dimensional signal detection model (2D-SDT). Different theoretical assumptions underlie threshold and signal detection model with debate in the literature which type of model better represents the cognitive processes in the task. As in Chapter 3, fitting both types of models allows me to compare performance of both models for this data set. If one type of model clearly outperforms the other, it gives an indication of which set of theoretical assumptions may better capture performance. The argument for fitting two different signal detection models follows the same logic. There is some debate whether separate memory dimensions need to be assumed to account for source recognition data, or if a model proposing memory performance on basis of a single memory dimension is sufficient to account for the data. In all three cases, I fitted a response bias model first. In this model, only the guessing and decision criteria parameters were fitted separately to actions not observed and actions observed five times.

The detailed results of the model fitting tests are reported in Appendix D.2.

Table 5.4. Parameter estimates for response bias 2HTM in Experiment 8

Observations	D_P	$D_R = D_N$	$d_P = d_R$	b	$a = g$
None	.73 [.67,.80]	.34 [.28,.39]	.74 [.67,.82]	.21 [.17,.25]	.21 [.15,.28]
Repeated	.73 [.67,.80]	.34 [.28,.39]	.74 [.67,.82]	.71 [.64,.77]	.45 [.38,.52]

Mean parameter estimates with 95% confidence interval in brackets

Two-High-Threshold model

The basic formulation of the 2HTM follows the one presented in Chapter 3. The assumption of multinomial tree processing models is that responding arises from certainty and uncertainty states. In certainty states, participants classify items based on correctly identifying the origin of the item. In uncertainty states, participants guess, with responses based on guessing leading to correct or incorrect identifications of items. The graphical representation of the model is in Figure 3.2 on page 116, with read actions replacing observed actions for this experiment.

In the 2HTM, the model is defined by the following parameters, all of which are defined as probabilities: item parameters D_P, D_R and D_N for performed, read and not presented actions, source parameters d_P and d_R for performed and read actions, item guessing parameter b and source guessing parameters a and g. All parameters can vary between 0 and 1. For item and source parameters, values greater than 0 present that responding is based on some memory strength for those items. For the guessing parameters, values of 0.5 represent unbiased guessing. For the item guessing parameter, values lower than 0.5 represent a tendency to respond conservatively and judge items as new, whereas values greater than 0.5 indicate a tendency to judge items as old. Values lower than 0.5 in the guessing parameter indicate a tendency to respond ‘Read’ (in this instantiation of the model) and values greater than 0.5 indicate a tendency to respond ‘Performed’. I manipulated the frequency with which participants observed actions in the second encoding phase. Some actions participants did not observe at all in the second encoding phase, while they observed other actions repeatedly. Both conditions are fitted simultaneously for all participants with parameters defined for both conditions separately.

I fitted the model using the *MPTinR* package (Singmann & Kellen, 2013). The equality restriction tests are described in Appendix D.2. The best-fitting model is the response bias model where only guessing parameters vary with observation; its parameter estimates are given in Table 5.4. The model shows that with observation participants are more likely

Table 5.5. Parameter estimates for response bias 2D-SDT model in Experiment 8

	D_P	D_R	d_P	d_R	c_{old}	$c_{performed}$
None	1.00 [.98,1.02]	.87 [.72,1.01]	1.00 [.97,1.03]	.00 [-.00,.00]	.90 [.81,.98]	.73 [.58,.87]
Repeated	1.00 [.98,1.02]	.87 [.72,1.01]	1.00 [.97,1.03]	.00 [-.00,.00]	.00 [-.00,.00]	.25 [.12,.38]

Mean parameter estimates with 95% confidence interval in brackets.

to respond ‘Old’ and that the ‘it had to be you [here: read]’ bias disappears when items are additionally encoded by observation. Participants’ source guessing is not biased towards the weaker source (as was the case in Experiment 4), and is the case in the baseline condition here.

Signal detection models

The basic formulation of the signal detection models follows the ones presented in Chapter 3 and 4. In a signal detection model, items are represented by (typically Gaussian) distributions on one or more memory strength dimensions. Participants make decisions by judging whether items’ strength exceed decision criteria on the dimensions. That means that participants’ decision is based on both items’ strength(s) and the placement of decision criteria. As in previous chapters, I will fit both the two-dimensional and one-dimensional signal detection model.

Two-dimensional signal detection model In the two-dimensional signal detection model, shown in Figure 3.3 on page 118, each source of memory is represented by a bivariate Gaussian distribution on separable but potentially correlated item memory strength and source memory strength dimensions. Participants judge items to be old (i.e., have been presented at encoding) if the strength of the item exceeds the c_{old} criterion on the item memory strength dimension. In this experiment, this means participants would judge an action to have been read or performed at encoding. Separately, if participants judge an item’s source strength to exceed the $c_{performed}$ criterion, they would respond ‘Performed’, otherwise they would respond ‘Read’. Changes in placement of the criteria reflect changes in participants’ response bias, while changes in placement or shape of the distributions reflect changes in the memory representations.

The parameters that define this model are item memory parameters for performed and read actions (D_P and D_R), source memory parameters for performed and read actions (d_P and d_R). Those parameters are set equal across observation conditions. The decision criterion c_{old} on the item memory and $c_{performed}$ on the source memory dimension vary with observation. I fitted the model using the *MPTinR* package (Singmann & Kellen, 2013).

Table 5.6. Parameter estimates for response bias 1d-SDT model in Experiment 8

Observations	D_P	D_R	c_{old}	$c_{performed}$
None	1.90 [1.74,2.06]	.75 [.60,.90]	1.03 [.94,1.14]	1.94 [1.81,2.08]
Repeated	1.90 [1.74,2.06]	.75 [.60,.90]	.05 [-.10,.19]	1.15 [1.00,1.30]

Mean parameter estimates with 95% confidence interval in brackets.

The results of the equality restriction tests are described in Appendix D.2. The response bias model that only allowed the decision criteria to vary with observation provided the best fit. The parameter estimates are given in Table 5.5. Observation shifted the decision criterion on both the item memory and source memory dimension to allow for more liberal responding on both dimensions (more 'Old' responses and more 'Performed' responses respectively), matching the model-free analysis.

One-dimensional signal detection model In the one-dimensional model shown in Figure 3.4 on page 120, each source of memory is represented by a Gaussian distribution on a single overall memory strength dimension. Participants judge items to be old (i.e., have been presented at encoding) if the strength of the item exceeds the c_{old} criterion. In this experiment, this means participants would judge an action to have been read or performed at encoding. If an item's strength additionally exceeds the $c_{performed}$ criterion, participants respond 'Performed' otherwise they respond 'Read'. Changes in placement of the criteria reflects changes in participants' response bias, while changes in placement or shape of the distributions reflect changes in the memory representations. As with the two-dimensional signal detection model, I tested whether the decision criteria (c_{old} and $c_{performed}$) would be affected by repeated observation, while the average strength of the distributions of performed and read actions would not (D_P and D_R respectively).

I fitted the model using the *MPTinR* package (Singmann & Kellen, 2013). The equality-restriction tests that identified the response bias model as the best-fitting instantiation of the one-dimensional signal detection model are described in Appendix D.2. The parameter estimates for this response bias model are given in Table 5.6. As with the two-dimensional signal detection model, the best-fitting one-dimensional signal detection model suggests that more liberal identification of items as 'Old' and 'Performed' accounts for the effects in the experiment.

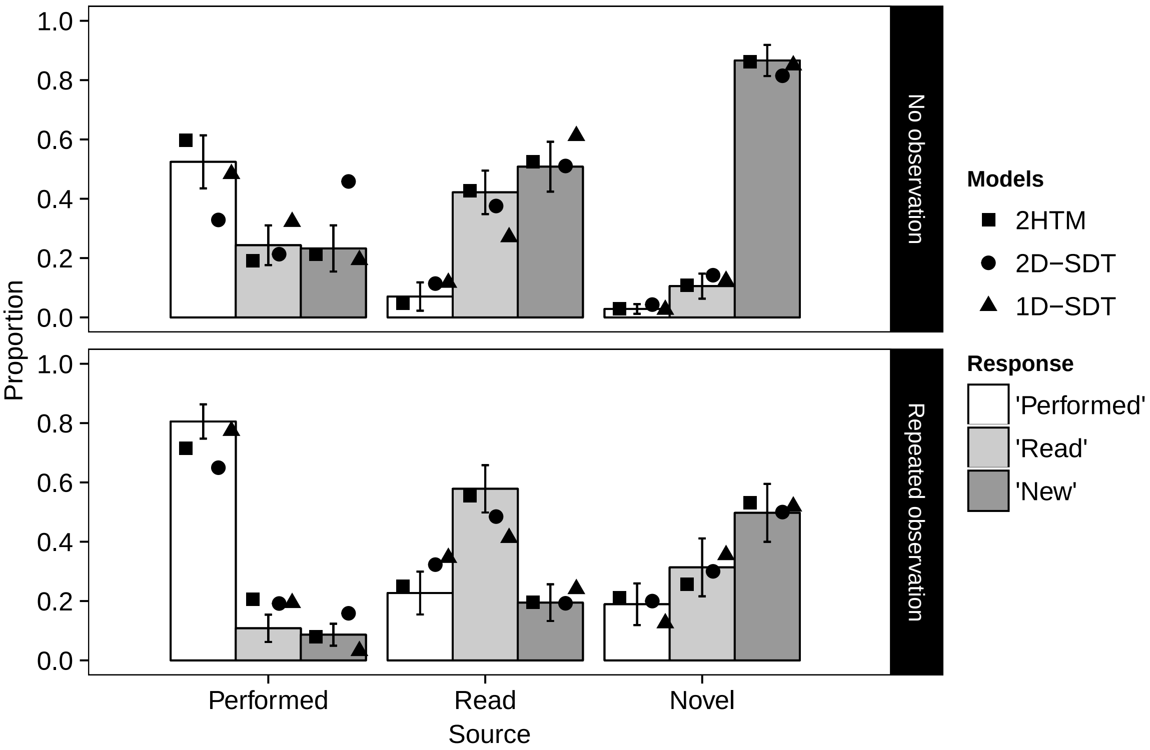

Figure 5.2. Proportions of responses by item type, observed and predicted by response bias instantiations of Two-Highthreshold model (2HTM), one-dimensional signal detection model (1D-SDT) and the two-dimensional signal detection model (2D-SDT). Error bars are 95% confidence interval.

Model comparison

Figure 5.2 shows the the relative proportion of ‘Performed’, ‘Read’ and ‘Not presented’ responses given to performed, read and novel actions respectively. For all three models, the data was fit to the raw frequencies of responses given in response to each source of item, separately for items not observed and repeatedly observed following encoding. The total number of items differed between sources. There were five performed actions, five read actions and five novel actions that were observed repeatedly. Yet while there were five performed and five read actions that were not observed, of the actions presented at test, 20 novel actions were presented as distractors at test that had not been observed. To conveniently plot the data points predicted by the model alongside the observed data, Figure 5.2 therefore shows the the relative proportion of ‘Performed’, ‘Read’ and ‘Not presented’ responses given to performed, read and novel actions respectively, rather than the raw frequencies. The figure shows that in fact, the 2HTM provides a good fit for the data, though the majority of data points for the 1D-SDT model equally fall within the 95% confidence interval of the observed data. The 2D-SDT model clearly shows some critical misfits.

Table 5.7. Model fits of response bias models in Experiment 8

		Aggregate				Individual					
Model	# P	G^2	p	AIC	BIC	G^2	p	AIC	AIC best	BIC	BIC best
2HTM	7	23	<.001	37	75	201	.19	719	20	2123	13
1D-SDT	6	77	<.001	89	121	305	.002	749	17	1951	24
2D-SDT	8	95	<.001	110	153	566	<.001	1158	0	2761	0

Table 5.7 shows the modelfits for the response bias instantiations for all three types of models for both the aggregate data and the sum of the individual data. In addition to the fits, the table also reports AIC and BIC and the number of participants best described by each model.

The model fits show that the two-dimensional signal detection model does not provide a good fit to the data compared to the other two models. While the Two-High-Threshold model fits better for aggregate and individual data than the one-dimensional signal detection model, it does so at the cost of an additional parameter. This leaves the 2HTM preferred over the 1D-SDT by AIC and BIC for the aggregate data. For the individual data, 1D-SDT is preferred by the BIC (but not the AIC). Both criteria penalize additional parameters but the BIC does so to a greater extent, leading it to prefer the least-parameter candidate model more than the AIC.

5.3.5 Discussion Model-based analysis

The model-free analysis showed strong effects of response bias when comprehensively analysing the data beyond the limited analysis typically performed on observation inflation data. Indeed, the fits of response bias models to the data suggests that participants pattern of responding could simply be the result of a shift in response bias after observation. Observation then does not appear to shift the shape or strength of the memory distributions but only affects the placement of criteria or guessing parameters (though see Witt, Taylor, Sugovic, and Wixted (2015) for an account of how response bias shifts can reflect shifts in perception, rather than only shifts in responding). Interestingly, if response bias drives responding, instructional manipulations should be able to reduce bias. Lindner et al. (2010) could not show any impact of instruction (warnings, examination of qualitative features indicative of performance) in reducing observation inflation. I will return to discuss the implications of this response bias account of observation inflation, after I discussed performance of the three different kinds of models fitted to the data.

There is debate in the literature whether threshold models or signal detection models (or more precisely the models' assumptions about the shape of the underlying memorial distributions) better represent the true memory distributions. As in Chapter 3, I therefore tested both classes of models to compare their fits to the data. The response bias instantiations of both types of models explained the data well. In particular both Two-High-Threshold model and Two-dimensional signal detection model model the same assumption of separable item memory and source memory dimension. The Two-High-Threshold model clearly outperformed the two-dimensional signal detection model here as it did for Experiment 4. The same caveats apply here that did in the previous analysis. My analysis severely restricted the two-dimensional signal detection model to be able to fit it at all, given the small number of available data points. This restricted representation may not make the best case for it. Second, the debate about Two-High-Threshold versus signal detection models largely concern their assumption about the shape of the underlying distributions (e.g. Batchelder & Alexander, 2013; Bröder & Schütz, 2009; Pazzaglia et al., 2013; Slotnick & Dodson, 2005; Yonelinas & Parks, 2007). I did not test for those here. If in fact the underlying memorial distributions are not represented by the assumptions of the Two-High-Threshold model, then a better fit to the data holds significantly less meaning.

The second point I discussed in the two previous chapters is contrasting the source monitoring account (Johnson et al., 1993), best represented by the 2D-SDT model or 2HTM, and the relative strength account of memory (Hoffman, 1997; Marsh & Bower, 1993), best represented by the 1D-SDT model. For this experiment, the two-dimensional signal detection model performed so badly that it was clearly outperformed by the one-dimensional signal detection model. In contrast to the 2HTM being favoured over both signal detection models in Experiment 4, the one-dimensional model competes on a more even level against the threshold model here.

While I focused on the aggregate data for much of the analysis, fitting signal detection models to aggregate data may skew results, as discussed previously (see A. L. Cohen et al., 2008; Lewandowsky & Farrell, 2011). Comparing which model (2HTM or 1D-SDT) better accounts for individual participant provides a fairer comparison then. Depending on information criterion, 2HTM or 1D-SDT better accounts for the data, with the BIC penalizing more strongly for the more complex model and favouring the 1D-SDT model while the AIC favours the 2HTM. This suggests that the relative strength account may be able to account for the data in this experiment (though future work should investigate individual fits more closely).

5.4 Discussion

This chapter had two aims. The first was to look at the effect of visual perspective on source misattributions. The predictions of both a mirror neuron network account of source misattributions after observation and a source monitoring account were that observing someone from a first-person perspective should increase false ‘performed’ responses. In fact, there was no evidence that observing someone from a first-person perspective over a third-person perspective changed participants’ source responses at test. This is in line with Lindner et al. (2010) showing that observing a video of an action being performed in a first-person perspective does not increase false ‘performed’ responses relative to to observing a video from a third-person perspective. It does not seem to be the case, as predicted, that the artifical nature of the video-presentation in particular prevented the predicted effect of perspective.

In contrast, Hornstein and Mulligan (2004) reported an effect of perspective. Importantly, in their experiment they changed the perspective of self-performance rather than the perspective of observation, with participants more likely to disown actions when they observed their actions in a mirror. It is possible that visual perspective may be used as a cue to identify observed actions but not performed actions (with performed actions also characterised by other features such as motor and somatosensory information). In that case, a fully crossed manipulation of visual perspective of performance and observation in the Hornstein and Mulligan (2004) would predict an interaction of perspective and encoding.

Given that both Lindner et al. (2010) and I report null effects of the visual perspective manipulation, and the remainder of source memory results on visual perspective are mixed, it is debatable if visual perspective acts as diagnostic characteristic to distinguish performed and observed actions. A more generic source monitoring account would predict that visual information in general is associated with external events (such as observed actions), while internally-generated events are not (Johnson et al., 1993). This is in line with the explanation Hornstein and Mulligan (2004) favoured that it is the quantity but not the quality of the visual information that led participants to increasingly disown actions. Interestingly, their data do not fit well with an even more general tenant of the source monitoring frame work that increased similarity of items should increase source monitoring failures overall (e.g. Johnson et al., 1988). In that case, source confusion for both performed and observed actions should have increased with increased similarity, but they showed that the increase was directional.

The second aim of this chapter was to more closely examine the mechanisms of the observation inflation effect (Lindner et al., 2010, 2016). The standard analysis reported for inflation experiments is the analysis of 'performed' responses (Lindner et al., 2010; Lindner & Henkel, 2015; Lindner et al., 2016, 2012; Schain et al., 2012) or 'imagined' responses (Goff & Roediger, 1998; Lindner & Echterhoff, 2015), with a focus on the difference of false responses to encoded and unencoded items. In this chapter, I performed this standard analysis but complemented it with the full source recognition analysis (as in Chapter 3) and model fitting.

The standard observation inflation analysis exactly replicates the pattern of data reported by Lindner et al. (2016) (with a different key experimental manipulation), indicating a successful replication of their paradigm. I separately looked at item memory and source memory for the full set of data as in Chapter 3. The pattern of item memory (increased identification of performed actions as 'old' with a decrease in d') matches the pattern of data reported by Goff and Roediger (1998). Similarly in the source memory analysis (of old items correctly recognized as 'old'), the increase in correct identification of performed actions came at the cost of a decrease in correct identification of read action phrases. The model-based analysis confirmed what this data suggests: a shift in response bias is critically responsible for the observation inflation effect observed in these experiments.

This means that observing someone else perform an action increases participants' tendency to respond 'performed' at test. Assuming that the mechanism underlying imagination inflation is identical, this suggests that imagination and observation (and possibly other visualiation manipulations (Henkel & Carbuto, 2008)) increase claims of self-performance, while other re-exposure manipulations (read, generate (Lindner et al., 2010)) do not.

The discussion then turns to the mechanism that makes observation and imagination so potent. It is possible that observation and imagination enrich the memory trace to such a degree that at test only responding 'performed' adequately matches the richness of the memory, with the limited response options given. Neither the 'read' nor the 'not presented' response options are characterised by memory features rich in perceptual detail, so are not viable candidates when classifying a perceptually rich memory. If simple enrichment is enough to cause the effect, this may explain why visual perspective (Lindner et al., 2010) or the degradation of the visual image (Lindner et al., 2016) do not additionally moderate inflation. The sheer presence of this enriching manipulation is sufficient to lead to inflation, while the details of the image are irrelevant.

This account would predict that if participants were given a different option of resolving the source information conflict, they may be able to avoid giving wrong responses. The impact test formats and instructions have on memory performance has been discussed extensively in the literature (e.g. Chan et al., 2012; Dodson & Johnson, 1993; Marsh & Hicks, 1998). In the current test format, participants are asked to indicate how they originally encoded an action event (performed, read, not presented) and to ignore the subsequent encoding by observation. Giving participants an additional option to indicate how they encoded an action in the second encoding phase (not observed, observed) gives participants an option to indicate a perceptually rich source that is not the wrong response. If it is the case that participants simply choose ‘performed’ for lack of an alternate option, this should occur less frequently in this changed format. If, on the other hand, participants choose ‘performed’ because motor system activation during observation provided evidence for the action having been self-performed, the additional option should not significantly reduce false responding. While participants should be less certain about their responses under the former account, they should be more certain under the latter account. A brief analysis of mean confidence ratings provided by participants (see Appendix D.3) confirms trends in that direction.

After the exploration of the motor trace in Chapter 2 and Chapter 3 and the cognitive trace in Chapter 4, this chapter looked at the specific influence visual perspective may have on source memory performance. I will move on to contextualise the empirical work in this thesis in the next, final chapter.

Chapter 6

General Discussion

In this thesis, I investigated how people are able to distinguish performed from observed actions. Eight experiments have focused on various features proposed to enable participants to distinguish performed from observed actions. Two theoretical accounts informed the present work: the motor simulation account proposed by Lindner et al. (2010) and the source monitoring account (Johnson et al., 1993). In this discussion, I will relate my experimental work to the broader research in the area. To integrate findings across experiments, I will not discuss the experiments in the chronological order of the thesis, but rather focus on integrating the contributions individual experiments make to the theoretical accounts. I will first address the experimental findings in my thesis that are relevant to the enactment effect (i.e. differences in item memory for performed and observed actions), before discussing the central aspect of my thesis: source memory for performed and observed actions. I will close on proposing some avenues of future work.

6.1 Enactment effect

When actions are encoded by enactment rather than verbally or by observation, participants typically show higher rates of retrieval of these actions: the enactment effect (for reviews see Engelkamp, 1998; Nilsson, 2000; Roediger III & Zaromb, 2010; Zimmer et al., 2001) .

I showed in Experiment 4 that a solid enactment effect emerges in a standard source recognition experiment, in which participants determine if a re-presented action was performed, observed or novel at test. Participants show better item identification for performed actions than for novel actions. This is in line with research in the enactment effect literature for Old/New recognition tasks (e.g. Engelkamp et al., 1994; Golly-Haring & Engelkamp, 2003; Mulligan & Hornstein, 2003).

I also observed significant enactment effects in the selective recall experiments I tested (Experiment 1, 2, 3, 5, 6 and 7). In particular the extended recall experiments allowed me to isolate availability of items at test (as a proxy for item memory). Overall, performed actions were more available at test, i.e. were more likely to be recalled than observed actions. While the enactment effect is typically more variable in free recall tasks (e.g. Engelkamp & Dehn, 2000; Helstrup, 2001; Schult et al., 2014; Steffens, 2007), my results are in line with within-set manipulations (participants take turns performing and observing actions) leading to robust enactment effects in recall (for a review see Steffens et al., 2015).

I observed the enactment effect for actions that were not verbally encoded (Experiment 1, 2 and 3), in line with Foley et al. (1991), and for actions participants performed with objects, in line with R. L. Cohen (1983).

Given the focus on source memory in this thesis, participants completed a selective recall task in contrast to the free report tasks typically used in the enactment literature (e.g. Engelkamp & Dehn, 2000; Golly-Haring & Engelkamp, 2003; Helstrup, 2001; Schult et al., 2014). When participants were cued to recall performed actions, they showed a robust enactment effect. When participants were cued to recall observed actions, on the other hand, the enactment effect shrunk considerably (see Appendix E for the data from Experiment 3, 5, 6 and 7). Similar relative improvements by constraining recall have been shown for verbal material (Hollins et al., 2016). While enacting an action may then improve item-specific processing by way of an encoded motor trace (e.g. Engelkamp et al., 1994) or integration the components of the action (e.g. Kormi-Nouri & Nilsson, 1998; Steffens, 2007), the retrieval context typically used in recall tasks in the enactment literature favours the enactment effect. It is not the case that observed actions are not encoded as well, but rather that they do not come to mind in competition with stronger, performed actions.

Throughout my experiments, I asked participants to generate the actions they performed. In contrast to findings by Nilsson and Cohen (1988) who failed to find evidence for generation enhancing enactment, I showed that when participants generate actions they perform they are later more likely to retrieve those actions than ones they only performed on instruction (Experiment 7). I used a within-subjects and within-sets manipulation of enactment (performed, observed) and generation (self-generated, partner-generated) while Nilsson and Cohen (1988) used a between-subjects design. Rather than the moderation-by-generation effect occurring due to increased item-specific processing of actions at encoding, the effect may only become apparent at retrieval in competition with weaker items in line with distinctiveness accounts (e.g. Hunt, 2012). If generation can moderate the size of the enactment effect, other manipulations may as well when tested within-subjects such as levels-of-processing (R. L. Cohen, 1981; Zimmer & Engelkamp, 1999) or enrichment (Helstrup, 1987; Nilsson & Cohen, 1988; Nilsson et al., 1995) which thus far have been argued to not moderate the enactment effect. This is also in line with the argument above that the enactment effect may not arise simply from encoding but also processes at retrieval.

Finally, in line with research by Steffens (2007), I observed that participants' memory of actions is enhanced by their memory for objects. In Experiment 4, I showed that false alarm rates increase dramatically when objects from encoding are re-used for distractor items (also see Engelkamp & Zimmer, 1997), compared to experiments that use unrelated distractors (e.g. Mulligan & Hornstein, 2003). Though encoded items and distractors would have been distinguishable by the unique combination of motor action and objects, participants' judgements appeared to be based on the identity of objects alone. If the motor trace is encoded separately alongside other memory traces of an item (e.g. Engelkamp, 2001), it appears to play a negligible role at retrieval.

6.2 Source memory in recall of actions

The studies of source memory of actions (specifically, source memory for performed and observed actions) in the literature have to date exclusively employed source recognition tasks at test. In this standard source recognition paradigm, participants encode actions by taking turns performing and observing actions. Participants complete the source recognition test immediately or after a delay. At test, the actions participants performed and observed, as well as novel actions, are read by or to participants. Participants are asked to determine for each item whether they performed the action, observed it or whether it is new (Foley et al., 1993; Hornstein & Mulligan, 2004; Leynes & Kakadia, 2013; Rosa & Gutchess, 2011). In these studies, participants' responding typically shows the following pattern: (1) better discrimination of performed from novel than observed from novel actions, in line with the enactment effect (e.g. Engelkamp et al., 1994), (2) a bias to attribute novel actions falsely identified as old to have been observed rather than self-performed, the 'it had to be you' bias (Bink et al., 1999; Hoffman, 1997; Johnson et al., 1993) and (3) failures in discriminating performed from observed actions, i.e. source misattributions

(Johnson et al., 1993).

To my knowledge, the present work is the first to look at source memory for actions in selective recall paradigms and to test if participants commit source errors when they are asked to recall through re-enactment performed or observed actions. I adapted the unconscious plagiarism paradigm (Brown & Murphy, 1989; Hollins et al., 2016) to use with actions. Participants took turns generating actions at encoding and at test were asked with cues (shapes in Experiments 1 through 3, sets of objects in Experiments 5 through 7) to recall (re-enact) actions they had performed or observed themselves (source-cued selective recall). When participants were asked to recall actions they had performed themselves, they correctly recalled actions they had performed themselves and falsely recalled actions they had observed or novel actions. The pattern of results for actions here mirrors the pattern of results in the recall phase of the unconscious plagiarism paradigm for verbal items (e.g. Brown & Murphy, 1989; Macrae et al., 1999; Marsh & Bower, 1993; Perfect et al., 2009) and therefore extends the unconscious plagiarism paradigm to actions. Importantly for the present work, it also shows that source errors for actions occur when participants' memory is tested with recall tasks.

As discussed above, typically source recognition of action studies show not only participants claiming observed actions as performed but performed actions as observed. In the present experiments, I therefore asked participants to recall not only performed but asked some of them to recall only observed actions. When participants were asked to selectively recall actions they had observed, they falsely reported some actions they had performed as observed. This replicates recent advances in research into unconscious plagiarism of verbal items. Hollins et al. (2016) have shown that if participants perform a selective source-cued recall of their partner's ideas, participants also commit this error and give away their own ideas to their partners (also see Hollins et al., 2015) in addition to misattributing partner's ideas to themselves. The present work thus extends this work to actions.

The pattern of results in the recall experiments mirrored the results produced by standard source recognition for action paradigms, in particular: (1) participants correctly recalled more performed than observed actions, replicating the enactment effect (e.g Golly-Haring & Engelkamp, 2003), (2) participants falsely recalled more novel actions as observed than novel actions as performed, replicating the 'it had to be you' bias (Bink et al., 1999;

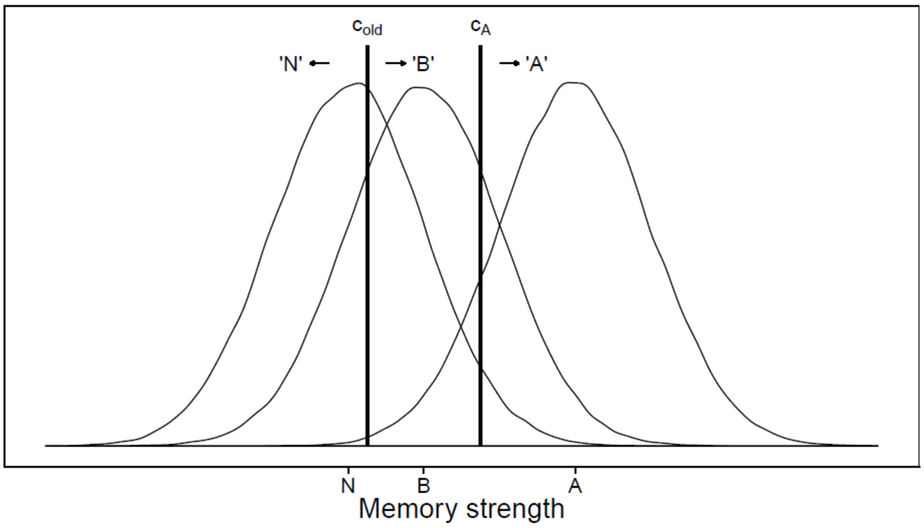

Figure 6.1. Illustration of relative strength account of source memory, with A the stronger, B the weaker source and N novel items, c_{old} the Old/New decision criterion and c_A the A/B decision criterion.

Hoffman, 1997; Johnson et al., 1993) and (3) participants showed failures to discriminate performed from observed actions by falsely retrieving the wrong-source actions during the selective recall task. Thus, source confusion of observed and performed actions occurs not only for autobiographical memories of childhood events (Ikier et al., 2003; Küntay, Gülgöz, & Tekcan, 2004), or when participants are asked to identify the source of a presented item (e.g. Foley et al., 1993; Hornstein & Mulligan, 2004; Leynes & Kakadia, 2013; Rosa & Gutchess, 2011), but also occurs when participants are asked to selectively access and report actions from only one source.

6.2.1 Comparison to source memory in recall of verbal items

Hollins et al. (2016) and Hollins et al. (2015) showed consistently that participants were more likely to give away their ideas to their partner than steal their partner's ideas. In the present experiments, some experiments showed no significant bias towards one over the other error (Experiments 1, 2, 3, and 7), while some did (Experiment 5 and 6). Even experiments that showed no significant bias towards participants being more likely to disown actions than claim them, still showed a trend towards it (with exception of Experiment 1). Biased source error recall could occur because of differences in memory strength between sources. Then (1) stronger items may intrude more frequently during the recall of the weaker source, (2) weaker items intrude less frequently during the recall of the stronger source or (3) a combination of both. This is in line with a memory strength account of source confusion (Hoffman, 1997; Marsh & Bower, 1993), see Figure 6.1.

Under this equal variance account, two changes could account for this bias in source errors.[1] First, the source criterion may be biased to liberal responding. However, examining the monitoring data in the extended recall tasks, that is the rate with which participants falsely report items from the wrong source as task-compliant, reveals no bias in Hollins et al. (Experiment 3, 1 week delay) with rates of .29 and .31 in the Recall own and Recall partner task respectively. For comparison, the rates I reported in comparable experiments in this thesis also do not show a clear bias in the monitoring data (with exception of Experiment 5).

Second, if Source A is signifcantly stronger than Source B, this alone may drive this bias under this account. Comparisons of the memory strength of sources in Hollins et al. (2016) and in the recall of action experiments I presented (see Appendix E) reveals a greater difference in memory strength between self-generated and heard ideas than performed and observed actions. This may account for the more consistent and larger bias in the verbal plagiarism literature than in the experiments I reported here.

An alternative explanation of the different patterns of data could be based not on the overall memory strength of items, but on differences in source memory strength (a source monitoring account (Johnson et al., 1993)). In other words, performed and observed actions may be more similar in their features than self-generated and heard items. Under such an account, the difference in response pattern should be evident in the monitoring phase of the extended recall task (if the assumption holds that this taps more or less exclusively into source memory, rather than item memory). It was not.

One hint as to the genesis of the bias may be in Experiment 2b in Hollins et al. (2016). Here participants recalled both own and partner-generated ideas at the same time (and sorted them into appropriate columns). The authors did not observe a bias to give away ideas in that experiment. Investigating whether this is due to changes in memory search, monitoring or reporting processes may shed some light on the difference between the (otherwise) consistent bias to disown verbal items but not actions.

Note that in Hollins et al. (2016), participants generated category exemplars such as ‘Fruits’ while they generated motor actions and interactions with objects in the experiments in this thesis. A better test of the differences between verbal and action task would require testing a verbal condition alongside the enacted conditions of Experiment 5 or 6, for example. In such an experiment participants would take turns generating actions. In

---

[1] Under assumptions of unequal variance, various combinations of a distribution with higher variance, differences in memory strength and placement of criteria could result in the same patterns of data

the verbal condition, participants would simply take turns saying the phrases out loud, while in the enacted conditions participants would take turns performing the phrases. In that case, the complexity of the items is equated and differences could be evaluated directly, all other things being equal.

One final note on using the relative memory strength account to attempt to explain these patterns of data is that this account struggles to explain data from recall experiments. As I showed in Chapter 4 and as Hollins et al. (2016) discussed, the high number of novel items that are recalled forces the distribution of novel items to be stronger than the distribution of the weaker source (observed actions or heard ideas) when fitting the data formally for the best fit. With novel items typically set to 0 for a baseline null-memory strength, practically this results in negative memory strength for encoded (observed or heard) items. I will discuss this in more detail in the section on the relative memory strength account below. It may be the case that in a recall, compared to a recognition task, participants sample far more potential intrusions and hence are more likely to report them.

6.2.2 Are source errors evidence for failures of source discrimination?

Errors in selective recall (more than errors in source recognition, though see Murnane and Bayen (1996)) can occur for reasons other than failures of source discrimination: (1) errors in recall tasks may simply reflect guessing and (2) errors in recall tasks may not only reflect source discrimination but also memory strength and retrieval processes.

Guessing

I addressed the first point in two ways. I asked participants to complete free report rather forced report tests. In forced report tests, participants are asked to recall the full number of encoded items, and to guess if necessary. In free report or free recall tests, participants are asked to simply recall as many items as come to mind and to step when they cannot remember any more items. Tenpenny et al. (1998) argued that forced recall increases source errors by making participants more prone to guessing.

Second, I developed an analytic approach that creates a baseline of expected guessing (or accidental source errors). I estimated what guessing performance would look like by simulating a baseline recall performance that was weighed by the typicality of items (i.e., as if participants were generating items at retrieval rather than recalling them from memory). For every participant, I was now able to compare the number of source errors they made in the experiment with the number of source errors I would expect them to make if they were simply guessing. I showed that guessing cannot account for all source errors participants reported, suggesting that source errors were at least in part genuine memory errors. This approach can be easily transferred to other recall tasks, such as the selective recall of verbal material. To create a similar baseline in Hollins et al. (2016), it would only be necessary to create a typicality profile of the generated category exemplars (or using existing norms (Van Overschelde, Rawson, & Dunlosky, 2004)), and code participants' data by the exact exemplars they generated at encoding and retrieved at recall.

The research on unconscious plagiarism has largely focused on average source error rates across items (for reviews see Gingerich & Sullivan, 2013; Perfect & Stark, 2008a), with some manipulations such a idea elaboration (e.g. Stark & Perfect, 2007) and source credibility (e.g. Perfect & Stark, 2008b) addressing the effect of differences in items and source that may drive plagiarism errors. To my knowledge, this guessing approach is novel in that it not only simulates guessing, but focuses attention on an item-level analysis of which items are likely to result in source errors. It is possible that the source errors reported in, for example, Hollins et al. (2016) (a) on average do not exceed the number expected by guessing and (b) are always the same items irrespective of which exact category exemplars participants generated. If those items are typical items, it is likely that guessing at least contributes to the rates of reported source errors. It would be possible to test part of this by asking participants to generate atypical exemplars. If participants report the same exemplars at retrieval that they report after generating typical exemplars, it would suggest that guessing does drive source memory failures in recall. Disentangling memory strength, typicality and source discrimination of typical and atypical items (whether they are verbal ideas or actions) may therefore benefit from an item-level analysis and guessing approach such as the one I suggested in this thesis.

I showed that guessing cannot account for the number of source errors I observed in the experiments. Further simulation and modelling work that was beyond the scope of this thesis should explore the boundary conditions of this guessing correction to test its validity. In the instantiation of the guessing simulation I presented, I assumed that all possible actions generated across participants are available to each participant at test as possible guesses. That is a very liberal assumption. Realistically only a subset of actions would be available to participants at retrieval, and the frequency of guesses would change accordingly. For example, in the present instantiation I assumed that all possible actions are available to participants at retrieval. Guessing, in that case, is based on the full distribution of actions and guessing rates are relatively low. However, to replicate the pattern of participants' responding exactly in a guessing simulation would require not only limiting the tail of the distribution but assuming that the distribution is different for each participant and includes only and exactly items they and their partner happened to have generated. This is unlikely.

Certainly one way to develop the guessing simulation is to look at the conditions under which performance can be approximated (and use it to explore the components that result in participants' performance at recall).

Limiting the overall size of the distribution (by cutting off its tail of less frequent exemplars) can lead to a better approximation of the number of source errors relative to the number of intrusion errors. A preliminary simulation of different tail lengths for a verbal unconscious plagiarism experiment suggested that with a distribution of 30% the size of the full distribution of exemplars, the ratio of source errors to intrusion errors can be replicated. Smaller distributions led to relatively more source errors, and longer distributions to relatively more novel items being guessed. This smaller distribution cannot account for the number of correct responses, even when some correct responses are set as being based on memory retrieval rather than guessing, and so guessing even on the basis of this smaller distribution cannot fully account for the pattern of data.

Another aspect worth exploring is participants' individual distribution versus an average or summed distribution across participants. If participants generated items at encoding, not only did they encode them memorially to some extent, they also showed which items they personally would generate (and likely re-generated). Weighting the typicality distribution by individual preferences (or memory strength) may then indicate that some items are more likely to re-appear at test than others. Modelling this boost, on top of the average typicality, could give an approximation of memory strength.

Memory strength versus monitoring

Source recall (that is correct and false source recall) in a standard selective recall task may be a combination of items with higher memory strength coming to mind more easily and being reported and participants' ability or failure to identify the source of an item correctly to report it. In other words, source errors may not only reflect source discrimination processes but also item memory.

To disentangle source discrimination from item memory, I used an extended selective recall task (Bousfield & Rosner, 1970; Hollins et al., 2016; Kahana et al., 2005) in addition to the standard selective recall task. The extended recall tasks separates the two aspects of report (according to generate-recognize model of recall (Watkins & Gardiner, 1979) into two indices of performance: availability and monitoring. The extended recall task showed that: (a) participants are more likely to report performed actions because performed actions are more available, in line with the enactment effect (Experiment 3, 5, 6 and 7), (b) when participants are asked to retrieve observed actions, this enactment effect is reduced (see Appendix 6.1 referenced earlier), (c) generating the idea for an action increases both the availability of those actions (Experiments 5, 6 and 7) and participants' ability to correctly monitor the source of the actions, if only only self-performed actions are generated (Experiments 5 and 6), but not if both self-performed and observed actions are generated (Experiment 7). In that sense, the monitoring aspect of the extended recall task seems to reflect source monitoring. The task also reflected that (d) source errors reduce when participants are encouraged to actively consider the source of their memory at retrieval (Experiment 3) and (e) concurrent load at encoding negatively affects the availability of actions when they are the target of retrieval and negatively affects monitoring when actions encoded under load are not the target of retrieval (Experiment 3). Finally, the extended recall task showed that (f) the 'it had to be you' bias, here the greater propensity for participants in the Recall observed than Recall performed task to report novel actions, arises at monitoring. It seems not to be case that more novel actions come to mind for participants when retrieving items from the 'weaker' source but that they are more likely to report them. This is in line with findings from Hollins et al. (2016).

Just as Kahana et al. (2005) and Hollins et al. (2016) argued, the extended recall task is sensitive to manipulations, in that both the item availability and monitoring aspect of the task are differentially affected in selective recall tasks. While the availability dimension seem to reflect item memory, the latter monitoring aspect of the task appears to reflect monitoring or late-correction processes (for similar approaches see Goldsmith, 2016; Guzel & Higham, 2013; Halamish, Goldsmith, & Jacoby, 2012). Further stochastic investigation of the dependency between item availability and monitoring indices as calculated in this thesis will be necessary. In this extended recall task, only items that are available to participants at test are reported or withheld. Similarly, in a standard source recognition task, source discrimination is only calculated for items correctly recognized as old. Murnane and Bayen (1996) showed that source discrimination is not free from influences of item recognition and guessing, so similar work is necessary for the extended recall task to fully understand the stochastic dependencies between measures.

Finally, throughout the thesis I used the monitoring measure of the extended recall task as shorthand for an index of source memory or source discrimination. Certainly, the effects of concurrent load (Experiment 3) and generating ideas for one of the to-be-discriminated sources (Experiment 5 and 6) in this extended recall task are in line with predictions of the source monitoring framework (Johnson et al., 1993), with the evaluation of source features of encoded items responsible for source memory performance. However, the monitoring measure itself does measure how participants make their monitoring decisions. While they may evaluate source features, participants' source memory performance may also be based on other aspects of the item.

I fitted two-dimensional and one-dimensional signal detection models to the data in Experiments 5 through 7. The dissociation of item availability and monitoring after item generation seemed to intuitively favour an account that dissociates item memory and source memory. Surprisingly, the one-dimensional signal detection model fared well in accommodating the dissociation in the availability and monitoring measures (with some caveats concerning the strength of novel items). Under this account, source memory performance is not based on a separate evaluation of source features but simply on an item's overall memory strength (Hoffman, 1997; Marsh & Bower, 1993). Testing the stochastic properties of the extended recall task (or testing it against predictions of the search of associate memory model (Raaijmakers & Shiffrin, 1980) as Kahana et al. (2005) suggested) were beyond the scope of this thesis but would shed some light on the information the extended recall task offers in understanding mechanisms of memory retrieval.

6.3 Theoretical accounts of source memory for actions

I will now turn to discuss the theoretical accounts I tested experimentally using these recall paradigms (in addition to standard source recognition and observation inflation paradigms) and by modelling the data. The most prominent account of how source decisions occur is a source monitoring account (Johnson et al., 1993) that argues that sources are inferred on the basis of qualitative features encoded alongside semantic item information. This theoretical account follows in tradition of multi-modal theories of memory (in analogy to multi-model accounts proposed to account for the enactment effect (Engelkamp, 2001)). An alternate, memory-strength based account claims that responding in source memory tasks may be accounted for entirely by items' overall memory strength, with source features not evaluated separately but simply subsumed into memory strength (Hoffman, 1997; Marsh & Bower, 1993).

More recently, Lindner et al. (2010) argued that false memories of performance after observation are the result of motor system activation during observation. In contrast to assumptions typically made by source monitoring accounts (Johnson et al., 1993; Leynes & Kakadia, 2013), this assumes that a motor trace exists not only for performed but also for observed actions. This motor simulation account only aims to explain the specific error of false memories of having performed an action after observation. In the original instantiation of their theory Lindner et al. argue that this presence of the motor trace leads to memories of observed actions to be falsely identified as performed and therefore leads participants to retrieve an observed action as self-performed. In the first half of my thesis, I tested the core claims of this theoretical account that the motor trace encoded during observation is crucial for false memories of self-performance.

6.3.1 Motor simulation account

The motor simulation account was proposed by Lindner et al. (2010) to explain why people claim that they performed actions when they only observed someone else perform the action. This account is based on neuroscientific evidence that observers' motor systems activate when they observe and action (Iacoboni, 2005; Rizzolatti & Craighero, 2004) and lead to creation of motor traces in the observer's motor system as if they had performed the action themselves (Grezes & Decety, 2001; Jeannerod, 2001). This account exclusively explains false claims of self-performance but cannot account for actions being disowned (though this has been shown in the literature, for example see Hornstein and Mulligan (2004)). The central prediction that follows from the motor simulation account is that observed actions will not be falsely identified as performed if the motor trace is not encoded during observation.

The original instantiation of this account was based on a series of null-results of source monitoring manipulations (Lindner et al., 2010). I tested its claims directly but found no evidence that limiting participants' motoric encoding of observed actions (Experiments 1 through 3) and the retrieval of the motor trace (Experiment 4) in dual-task experiments prevented false retrieval of observed actions as performed. There was no evidence for a reduction in the source error, when I occupied participants' motor systems with motor movements (Experiment 2 and 4) or with action plans (Experiment 2 and 3).

Lindner et al. (2016) recently claimed that incompatible motor movement concurrent to observation leads to a reduction in the false retrieval of observed actions as performed. While formally correct, closer examination of their experiment reveals that this conclusion is premature. They asked participants to perform motor movements incongruent to the motor action component of the action phrase they observed being performed with an object. They assessed the rate of false and correct performed responses after observation. Rather than comparing their data to a baseline condition of no motor movement, they compared it to a condition of congruent motor movement. In that condition participants performed movements congruent with the motor action component with the action phrase they observed being performed with an object. This condition led to an increase in correct 'performed' responses and false 'performed' responses relative to the incongruent condition. Critically, in this condition participants effectively pantomimed the actions along with observing someone perform the action. A false 'performed' response after observation now is participants responding 'performed' to an action they merely read before when they also pantomimed the action at a later point. Rather than incongruent movements reducing observation inflation, the results could just as easily be explained as congruent movements increasing observation inflation.

Given Lindner et al. (2016) observed a significant effect of their experimental manipulation there [2], simply testing the control condition against the incongruent movement condition would be a better test of the motor simulation account. If motor simulation is critically responsible for the effect false retrieval of observed actions as performed, performing incongruent movements during observation should decrease observation inflation relative to not performing movements during observation.

One possibility is that motor interference tasks used in this thesis and by Lindner et al. (2016) (a) do not affect motor simulation in the observer's motor system, or (b) do not affect memory of motor actions. For (a), observing motor actions (arm or finger movements) influences online concurrent execution of actions in direction of the observed action (Kil-

---

[2] Note that the significant effect of the manipulation was based on a joint effect on correct and false 'performed' responses. This deviates from their original approach of looking only at false 'performed' responses. Note also that for the critical comparison, two congruent movement (between-subjects) conditions were collapsed to compare against a single incongruent movement condition.

ner et al., 2003; Press et al., 2011). This suggests that they can lead to motor system activation during observation and there is little reason to believe that they did not in the present experiments. For (b), memory for performed actions decreases when participants perform an additional motor movement (walking, hand movement) immediately following the target action (Helstrup, 2001; Saltz & Donnenwerth-Nolan, 1981; Smyth et al., 1988), Similarly I showed in Experiments 2 and 3 that concurrent motor activation during observation reduces participants' retrieval of observed actions, but does not increase their source memory performance.

A second aspect of my experiments additionally calls the necessity of the motor simulation account into question. The motor simulation account focuses on the false retrieval of observed actions as performed. I showed that participants also falsely retrieved performed actions as observed (Experiments 1 through 7, with the design in Experiment 8 not allowing for the error). This is in line with other research into source memory for actions using the standard source recognition paradigm (Hornstein & Mulligan, 2004; Leynes & Kakadia, 2013; Rosa & Gutchess, 2011) and in line with source memory for verbal material (Hollins et al., 2016, 2015). The motor simulation account does not predict this error (nor claims to account for it). If motor simulation is uniquely responsible for false memories of performance, this would necessitate a separate explanatory account for false memories of observation. Alternatively, source memory accounts such as the source monitoring framework (Johnson et al., 1993) or memory strength based accounts (Marsh & Bower, 1993) can account for both errors, irrespective of the items' modality.

Finally, I examined the methodological approach that Lindner et al. (2010) based their motor simulation account on. In the observation inflation paradigm participants are asked to discriminate performed from read and novel actions after observation. Closer examination of the data using extensive source recognition and model-based analyses showed that the pattern of data typically reported in observation inflation experiments can be accounted for entirely by shifts in guessing (in threshold models) or response bias (in signal detection models). This suggests that observation does not decrease participants' ability to discriminate performed from observed actions, but simply increases their tendency to respond 'performed' at test.

The question then becomes why observation increases participants tendency to respond 'performed' while mere re-exposure to items does not (Lindner et al., 2010). Thomas et al. (2003) suggested that the sensory enrichment by imagination gives rise to the imagination inflation effect where items imagined after initial encoding are more likely to be identified as ‘imagined’. In analogy, the same mechanism could be responsible for the observation inflation effect (though as shown above, it cannot be just the presence of the motor trace). This would suggest that participants are failing to discriminate performed from now de-facto observed actions on the basis of the overlap of qualitative features (a source monitoring account explanation). This, however, assumes that participants’ evaluation of source features guides their source responding. I will discuss the limited evidence for that below.

A second aspect may contribute to the shift in response bias is the response choices participants are given. Participants in the observation inflation paradigm are never given the chance to make a correct source judgement for actions they observed: there is no option to respond ‘Observed’ in the source recognition test. Responding ‘performed’ may simply be the least-wrong response option. Given the influence of test format and instructions on performance in memory tests (e.g. Dodson & Johnson, 1993; Marsh & Hicks, 1998; McCloskey & Zaragoza, 1985; Mulligan, Besken, & Peterson, 2010), future work on the observation inflation paradigm should therefore include manipulations of the responses that are elicited from participants and how they are elicited. While it is possible that observation decreases discrimination of performed and previously read (but now observed) actions, the current test format does not allow examination of it and may additionally bias responding.

6.3.2 Source monitoring account

As I indicated, the direction of results when testing the motor simulation account pointed towards the data being better explain by the source monitoring account. Under this account qualitative features are encoded alongside semantic item information, and the source of an item is inferred on the basis of these features (Johnson et al., 1993). Accordingly, different qualitative features are diagnostic of different sources, with, for example, features for performed and observed actions overlapping more than features for performed actions and read items would (Leynes & Kakadia, 2013).

The majority of the experimental results I reported can potentially be accounted for within the source monitoring framework. I showed that interference (both motor and verbal interference) increases the number of source errors participants report relative to baseline.

In fact, this increase in source errors seems to be due to participants being less able to discriminate performed from observed actions if they observed actions under distraction. This is in line with Jacoby, Woloshyn, and Kelley (1989) reporting that recollection of details suffered after encoding under divided attention and Lane (2006) showing that participants are more prone to the misinformation effect when they encoded the original event under divided attention. Similarly, Macrae et al. (1999) showed that when participants took turns generating verbal items under auditory distraction, they were more likely to subsequently plagiarise their partner's ideas at test.

The source monitoring framework also predicts that interference at retrieval should limit participants' ability to correctly retrieve the source of their memories (Johnson et al., 1993). I tested whether motor or visual interference at test would affect source recognition performance in Experiment 4. There was no evidence for a detrimental effect of this low-level interference. Indeed, it seems to be the case that interference at retrieval affects the processing or monitoring of information at test, rather than the retrieval of specific features (e.g. Craik et al., 1996). Dodson et al. (1998) reported that source memory for specific source features suffered when participants tried to recollect information while engaged in a digit monitoring task. Similarly, Troyer and Craik (2000) showed that a digit monitoring task at retrieval disrupted retrieval of item and source features (also see Craik, Luo, & Sakuta, 2010). This suggests that more cognitively demanding interference tasks would have disrupted source retrieval, likely via disruption of the processing of source information. As it stands, there is no evidence that the specific modular (motor, visual) interference can interfere selectively with the retrieval of those same features of participants' memories. This calls into question how important individual source features are in making source decisions, and by extension, calls into question what participants are evaluating when making source judgements.

The flipside to impaired processing of information at retrieval, is encouraging participants to increase their monitoring of source information at retrieval. Marsh et al. (1997) showed that when participants were asked to carefully consider the source of the verbal items they recalled, they were less likely to plagiarise their partners. In line with this, I showed in Experiment 3 that if participants are asked to carefully monitor the sources of items they retrieve, their discrimination of performed and observed actions improves (if they sufficiently encoded the items). Thomas and Bulevich (2006) showed similar effects in the imagination inflation paradigm. When they asked participants to focus on the qualitative features of their memories, participants were less likely to respond ‘Performed’ rather than ‘Read’ or ‘Novel’ in the imagination inflation paradigm. In contrast, Lindner et al. (2010) failed to replicate the effect in the observation inflation paradigm (though note, given one study showed and the other did not show the effect, the evidence is far from conclusive). It is possible that Thomas and Bulevich presented the simpler task. In line with predictions from the source monitoring framework, performed and previously-read-now-imagined actions are less perceptually similar than performed and previously-read-now-observed actions (Hashtroudi et al., 1990; Johnson & Raye, 1981; Lampinen et al., 2003). Asking participants to focus on the differing qualitative features would be more helpful in the former than the latter case.

I used the framework suggested by Leynes and Kakadia (2013) to guide the experimental manipulations of individual features for the experiments in this thesis. According to the framework, performed actions are characterized by the presence of a motor trace and cognitive trace. In contrast, observed actions do not contain a cognitive trace (and the presence of the motor trace is debatable, as shown above). I argued above that if observed actions are enriched with a motor trace via motor simulation in the observer’s motor system, this motor trace does not appear to be responsible for false memories of self-performance. However, there is a limitation to that claim. In Experiments 1 to 4 in which I tested the possible impact of motor encoding of observed actions, performed and observed actions differed in more than the presence of a regular cognitive trace associated with self-initiated actions. Participants in those experiments actually generated the actions they performed themselves. That is, whenever participants performed an action themselves, they had generated the idea for that action. Whenever they observed someone else perform an action, they not only were not executing the action but they also had not come up with the idea for that action.

I tested whether generating the idea for an action would enhance the cognitive trace associated with performed actions and make it easier for participants to discriminate performed and observed actions. Indeed, I showed in two experiments (Experiment 5 and 6) that participants were more likely to correctly recall performed actions and less likely to falsely recall performed actions as observed when they had generated the idea for the performed action. This provided preliminary evidence in favour of the account proposed by Leynes and Kakadia (2013) that performed, but not observed, actions are characterized by a cognitive trace. However, the direct test of this diagnostic features account failed to find positive evidence. Participants did not blindly identify generated actions as performed. When participants generated both performed and observed actions (Experiment 7), participants were more likely to recall generated actions, but were not more likely to claim observed actions they had generated as performed than observed actions they had not generated. This suggests that participants weigh whether available source features are diagnostic of a source (indicate one but not the other source) at retrieval, rather than simply decide whether a feature associated with a source is generally diagnostic of a particular source. Source features therefore seem to be evaluated heuristically.

Finally, I looked at whether visual perspective would similarly clearly define performed actions (first-person perspective) and observed actions (third-person perspective), as suggested by Leynes and Kakadia (2013). I tested the claims in an observation inflation paradigm. I found no evidence that manipulating participants to observe actions from a first-person perspective (Experiment 8), increased their tendency to claim those observed action as self-performed. This replicated findings from Lindner et al. (2010) who similarly showed that moderation of the observation perspective does not modulate participants' tendency to claim observed or imagined actions as performed. In contrast, Hornstein and Mulligan (2004) showed that participants were more likely to misattribute performed actions as observed when they observed themselves from a third-person perspective. Lindner and Echterhoff (2015) reported that when participants observed themselves from a third-person perspective, imagining someone else perform an action increased false 'performed' responses relative to when they had simply performed actions without observing themselves. Interestingly, they also reported that self-imagination led to more claims of self-performance than other-imagination (not significant), while other-imagination led to more claims of other-performance than self-imagination (significant). Though note, that this is an imagination and not an observation manipulation.

It is possible that the mechanism underlying false 'performed' responses in the observation inflation paradigm limits the effects of any source feature manipulation. If observed actions are identified as performed simply on the basis of an overall perceptually rich trace, the specific details of that representation may not additionally moderate the effect. Regardless of a first-person or third-person visual perspective, this enriched memory will be identified more frequently as a performed than a read or novel action. This would also explain by the visual richness of the observed action did not moderate the observation inflation effect in Lindner et al. (2016).

A second possibility is that the theoretical framework proposed by Leynes and Kakadia (2013) is wrong on this account. Rather than a third-person visual perspective being diagnostic of observed and first-person visual perspective being diagnostic of performed actions, observed actions may simply be richer in visual information in general and performed actions poorer (Lampinen et al., 2003; Mitchell & Johnson, 2000). This view predicts the results by Hornstein and Mulligan (2004) who only observed participants increasingly disowning actions with increasing visual richness of performed actions, rather than commit more source errors overall with increasing similarity of sources (as may be expected e.g. Dodson et al., 1998; Hollins et al., 2015; Macrae et al., 1999). This interpretation of the source monitoring framework would also explain why visually rich observed actions are not more likely to be misattributed as performed (Lindner et al., 2016).

Leynes and Kakadia (2013) proposed that the specific visual perspective is diagnostic of performed and observed actions. I found little evidence for this view. Leynes and Kakadia argued that performed, but not observed, actions are characterized by a cognitive trace. While I showed that adding a cognitive trace to one but not the other source increases discrimination, I found no evidence for this being due to performed actions in particular being associated with that trace. Additional proposals concerned the motor trace and the somatosensory trace as diagnostic characteristics of performed actions in particular. Some ideas of how to test those aspects of their framework will be presented below.

The alternative to the diagnostic features account is a broader, more general source monitoring account that is based on the notion of overall similarity of items increasing source errors. The results I presented are all roughly compatible with that account. Arguably, there is little experimental evidence that would not be compatible with some interpretation of the source monitoring framework given its flexibility. This means that while the source monitoring framework is great in guiding experimental exploration, it is difficult to directly test predictions and more difficult to falsify (note, it is a framework, not a theory). The results I presented in the thesis are roughly compatible with the source monitoring framework in general, but only partially compatible with a specific instantiation of it that proposes that sources are a priori defined by qualitative features that guide source decisions.

6.3.3 Relative strength account

Typically the source monitoring framework proposes that source features are evaluated at retrieval to make source judgements (Johnson et al., 1993). While participants in my experiments monitored the source of their memories more accurately than not, manipulations of individual features did not clearly indicate that specific features at test had an isolated impact on source decisions. One interpretation of this account is to assume that source features collapse into a composite source memory strength distribution, while the semantic item memory is represented by a separate item memory strength distribution. Mathematically this assumption can be represented by a two-dimensional signal detection model or a Two-High-Threshold model of source monitoring (e.g. Banks, 2000; Bayen et al., 1996; DeCarlo, 2003; Hautus et al., 2008; Slotnick & Dodson, 2005; Wickens, 2011)

A competing view is that not only features collapse into a source strength dimension but that source monitoring judgements are made based on a composite memory strength estimation that subsumes features and semantic memory strength (e.g. Hoffman, 1997; Marsh & Bower, 1993). This is best represented by a one-dimensional signal detection model, and assumes that memory judgements of any kind are simply based on the relative strengths of memory distributions.

I fitted both threshold and signal detection models to experiments using source recognition (Experiment 4) and observation inflation paradigms (Experiment 8). Here, threshold models outperformed signal detection models in overall fit. Threshold models assume that the distributions of the underlying memory representations are rectangular (items are recognized or not recognized), whereas signal detection models assume a continuous distribution of memory strength for items. The good fit of threshold models has also been shown by Schütz and Bröder (2011) for confidence rating data where the guessing parameters in threshold models and the bias parameters in signal detection models were allowed to vary for the different levels of confidence.

It was beyond the scope of this thesis to test why Two-High-Threshold models outperformed signal detection models. One possibility is that they are more flexible in accommodating data. One way to test this in future work is by not only fitting models, but also fitting data predicted by individual models by predicting and competing models in model recovery procedures. If models are too flexible in accommodating data, they will not only be recovered as the correct model for the data they predicted but also as the preferred model for data generated by the competing model (see Wagenmakers, Ratcliff, Gomez, & Iverson, 2004, for a description of this procedure).

Second, while threshold models provided a good fit to the data, I did not test whether my data reflected the underlying memorial assumptions of either threshold or signal detection models. Proponents of threshold models argue that threshold models constitute a solid analytical method that addresses some of the pitfalls of standard analyses, such as analysing source memory measures contingent on item recognition (Batchelder & Alexander, 2013). In contrast, proponents of signal detection models argue that the assumptions of thresholds do not represent the true nature of memory distributions. They show that when participants are asked to give confidence ratings (as a proxy for assessing differing memory strengths for items of one distribution), ROC curves follow the signal detection prediction of curvilinear, rather than the threshold model prediction of straight, curves (e.g. Slotnick & Dodson, 2005; Slotnick, Jeye, & Dodson, 2016). That analysis was beyond the scope and original focus of this thesis and would require collecting and analysing confidence ratings. I collected confidence ratings for the final experiment in this thesis (Experiment 8). The observation inflation design in that experiment relies on a very low number of items per item type, so analysing the data by constructing ROC curves to test their curvature would have been noisy on the level of an individual or an aggregate analysis.

An alternate account to the source monitoring framework that argues for a separate evaluation of source features at test (Johnson et al., 1993), is the relative strength account of source memory performance (Marsh & Bower, 1993). Contrary to the assumption of two memory dimensions, this account assumes only a single dimension of memory strength that accounts for both item memory and source memory performance. This account is in line with more recent arguments of a single memory distribution underlying performance in a variety of tasks (e.g. Berry et al., 2008; Shanks & Berry, 2012).

Teasing apart the relative strength account and the source monitoring account experimentally is difficult since participants' evaluation of items' source and resulting monitoring performance could be based on strength or source features. As I showed above, manipulating source features directly that I did not expect to result in changes in items' memory strength (Experiment 4, Experiment 7, Experiment 8), did not result in changes in participants' source memory performance. When I manipulated source features that additionally affected the memory strength of items (Experiment 3, 5 and 6), participants' source memory performance changed. This provides some evidence for a memory strength account over a source monitoring account.

I also compared both theoretical accounts formally by testing the two-dimensional signal detection and 2HTM model against the one-dimensional signal detection model, with the former representing the more complex source monitoring account (Johnson et al., 1993) and the latter representing the relative strength account (Marsh & Bower, 1993).

I compared the one-dimensional and two-dimensional model in the source recognition (Experiment 4) and observation inflation experiment (Experiment 8), and additionally in some of the recall experiments (Experiments 5, 6 and 7). The relative strength account (represented by the one-dimensional model) fares well in accounting for the data compared to the other two, more complex accounts, despite its relative simplicity. It was preferred over the two-dimensional signal detection model in Experiment 4 and 8, at least equal to the Two-High-Threshold model in Experiment 8 and accounted for the seeming dissociation of item and source memory in the recall experiments (Experiment 5, 6 and 7).

The relative strength account struggled the most in accounting for data in the source-cued recall experiments. While the model did generally well predicting responding for performed actions and for novel actions, the one-dimensional model struggled to accommodate the responding to observed actions. In a recall task, participants tend to produce a high number of intrusion errors (more than they produce false alarms in a recognition task). A relative strength account of memory needs to be able to explain why, when participants recall items from the stronger source, novel items that were not encoded are more likely than items from the weaker source to be reported. At the same time, when participants attempt to recall items from the weaker source, they are able to do so and do not simply produce novel items.

A model that assumes a separate dimension for analysis of source features can account for this pattern of data easily by assuming that the evaluation of source features contributes to performance in a source-cued recall task. A relative strength model has to accommodate the pattern in a different way. One way (shown in Experiments 5, 6 and 7) is to assume that the average memory strength of novel items falls between performed and observed actions (or in other words, that observed actions have negative memory strength if novel actions are defined with 0 memory strength). This seems counter-intuitive, but the same issue has been discussed in Hollins et al. (2016), so it is not a consequence of, for example, different ratios of memory strength of actions and verbal items.

It may be necessary to re-examine what memory strength means in a recall (versus a recognition) task. For example, Marsh and Bower (1993) labelled their memory strength axis in the proposal of this relative strength account as ‘associative strength’. Under those assumptions, representing observed actions with a negative associative strength in a recall task assumes that novel items generated at retrieval are more closely associated with performed items than observed actions generated by another person. Given that participants in a recall task have to recall items, rather than just recognize them, some measure of accessibility or cue-ability or association that leads participants to recall items may be necessary to explain memory in recall tasks. Simply adapting the signal detection models from recognition provides a good starting point to account for the data, but combining the relative strength approach and an associate approach like SAM (Raaijmakers & Shiffrin, 1980) by adding an additional parameter to the recognition model may be a more successful way forward.

As it stands, my thesis provides some evidence that participants may in fact primarily distinguish performed from observed actions by their differing memory strengths alone.

6.4 Future work

I have already outlined two areas of further investigation. The first one is to validate the guessing simulation and test it for non-action data (discussed earlier in the discussion), to explore source memory performance by focusing on an item-level analysis of which ideas or items lead to source errors. The second area is the inflation paradigm to investigate the mechanisms underlying the effect. In the discussion of Chapter 5, I suggested that participants’ false responses may be the result of not having an option to give an ‘Observed’ response. A simple change to the test format would allowing testing that account.

Beyond those strands, there are two additional areas of research that follow the experimental work in this thesis.

6.4.1 Enactment effect

The focus on my thesis was not to explore mechanisms of the enactment effect. Nevertheless, enactment research informed some of the work here and my results contribute to the enactment literature. There are two strands that I believe would be worthwhile to pursue.

First, I showed that the enactment effect is reduced in size and even eliminated when participants are focusing their recall on observed, rather than performed actions. Those findings are in line with some of the variable enactment effect results in free recall (Steffens et al., 2015). This suggest that retrieval mechanisms contribute to the enactment effect in addition to simply a stronger memory trace following encoding by enactment.

Future research here should focus on what makes performed actions more easily accessible. It is possible that the higher memory strength allows them to ‘win’ in direct competition against observed actions. The alternative option is that it is not memory strength per se but that they are more easily recovered or accessed in memory. One angle to explore this would be to ask participants to selectively recall performed and observed actions or recall both performed and observed actions. Rather than focusing on quantity of recall, it may be useful to investigate which performed and observed actions come to mind in which condition (to compare relative contributions of memory strength on an item-level) and how items are organized at recall to understand participants search patterns of memory and how they may contribute (extended work from Golly-Haring & Engelkamp, 2003).

The second aspect worth exploring are the manipulations of increased processing of enacted items that have been reported as failures in the enactment effect literature thus far (R. L. Cohen, 1981, 1983; Helstrup, 1987; Nilsson & Cohen, 1988; Nilsson et al., 1995; Zimmer & Engelkamp, 1999). I showed that generating actions can lead to an increased enactment effect relative to not generating actions in a within-set, within-subjects manipulation. Since the studies investigating effects of processing are between-subjects manipulations of processing, it is possible that simple running the same studies with within-set manipulations would result in the predicted effects.

6.4.2 Source features versus memory strength

In this thesis I used the diagnostic features framework suggested by Leynes and Kakadia (2013) to guide the experimental manipulations. Based on the source monitoring framework (Johnson et al., 1993), Leynes and Kakadia suggested that performed actions are characterized by a cognitive trace, motor trace, first-person visual perspective and a somatosensory/proprioceptive trace. Observed actions in contrast are characterized by the third-person visual perspective.

Under that account, there are two features I have not explored in this thesis: the motor trace encoded alongside performed actions and the proprioceptive features encoded when performing actions. Under the diagnostic features account, removing those features from performed actions should reduce identification of performed actions, adding them to observed actions should increase false identification of observed actions as performed, with either manipulation reducing participants' ability to discriminate performed and observed actions overall.

Manipulating the motor trace encoded when performing actions is difficult. It would require having participants perform distracting actions while performing their target actions. There are, in principle, three ways of achieving it. One would be to ask participants to perform the distracting actions just following the target action (assuming retroactive motor interference) to disrupt consolidation of the target motor memory (Krakauer & Shadmehr, 2006). This has been used in the enactment literature with some success in reducing the enactment effect (e.g. Saltz & Donnenwerth-Nolan, 1981; Zimmer & Engelkamp, 1985). The difficulty in realizing the experiment would be in equating conditions. Participants would have to perform actions with retroactive motor interference and observe actions with retroactive motor interference. Naturally, this would require an additional interference manipulation to test that it is not just the presence of interference alone. It may be possible to test motor against visual interference, following performed and observed actions (similar in spirit to Saltz and Donnenwerth-Nolan (1981)). While motor interference should disrupt identification of performed but not observed actions, visual interference should do the opposite (though as a consequence both should disrupt source discrimination). It would be possible to test this using the motor actions from the first three experiments in this thesis, to minimize the influence of memory for objects in cueing identification.

A second approach would be to ask participants to perform concurrent movements with a part of the body they are not using to perform the action, i.e. to move their legs while they perform an action with their arms. This approach assumes that any motor activation would disrupt the motor information resulting from performing the target action. Since motor actions for different body parts are encoded in different parts of the motor cortex, it is not clear whether that would lead to disruption of the initial motor memory or would not, more likely, lead to simply encoding the whole action complex as the target action

A third approach would be based on mirror neuron activation. Accordingly, participants encode actions they observe motorically (Iacoboni, 2005) and concurrent perception of actions influence online action execution (Kilner et al., 2003). In a complement to the approach in Experiments 2 and 3, observing someone else perform a distracting action while performing a target action oneself, should disrupt encoding the motor component of that target action. Simply adapting the design in Experiment 2 and 3 should allow testing this account, with the prediction that source discrimination will be reduced if participants observe simple actions being performed as they perform actions.

A related experiment of that paradigm in particular should test the relative impact of visual information. It is possible that participants source memory performance is not only mediated by information of who generated the action but may be mediated by visual information. Since it is not possible to eliminate visual information when observing someone perform an action, any disruption of correct identification of performed actions (and related increase in source errors when recalling observed actions) may simply be the result of visual interference if participants simply visualize what they look like when they perform actions. A possible experimental exploration of this would involve non-motor visual interference such as the sequence of faces. If participants encode actions they perform visually, watching images or sequences of images, should disrupt memory for that component.

The final trace that Leynes and Kakadia (2013) proposed that allows distinction of performed and observed actions is the proprioceptive or somatosensory trace, i.e. the feedback from performance of actions that is absent when observing actions being performed. Two possible approaches to test the importance of this trace are as follows.

In the first approach, I would adapt the motor action paradigm I used in Experiments 1, 2 and 3. Here participants would be asked to perform actions wearing weights on their hands and feet to increase the effort/afferent feedback from any motion. Compared to a condition where participants perform actions without weights, participants should be better able to discriminate performed and observed actions at retrieval. An additional manipulation would be one of matching feedback at retrieval and encoding. In line with the encoding specificity principle (Tulving & Thomson, 1973), participants' source performance should be higher for matching than for non-matching conditions.

In a similar experiment, a paradigm asking participants to perform actions with objects could be modified for this manipulation. Rather than attaching weights to participants (which may result in changes in performance based on that visual cue, for example), one possibility would be to change the weight of the object. Keeping all other things equal (size, shape and color of the object, such as using coloured boxes), filling the boxes with lighter or heavier material would change the feedback participants retrieve. In a recognition test, I would expect participants to be more likely to correctly identify an action as performed if the object has the same weight, than if the object has a different weight.

The second approach focuses on the action intention. In an adaptation of an action execution and perception experiment from Zwickel, Grosjean, and Prinz (2010), I propose manipulating whether participants perform actions themselves or are guided by someone to perform an action. While the proprioceptive feedback participants receive would be the same if they chose and made two objects interact or if someone else guided their hands to make two objects interact, the motoric planning would differ. This approach may allow distinguishing the relative impact of motor and proprioceptive information which are typically confounded.

Note that these experiments focus on manipulating source features without manipulating or while controlling for items' memory strength. The complement to this is to manipulate the overall memory strength of items without manipulating source features.

Items' memory strength decreases with delay or if items are encoded under distraction (e.g. Jacoby, Woloshyn, & Kelley, 1989) and increases with repetition or deeper processing (e.g. Lockhart & Craik, 1990), for example. While those manipulations do not overtly manipulate source features being added to one or the other source, they likely would strengthen or limit encoding of source features.

In a levels-of-processing experiment it may be possible to manipulate the processing for performed and observed actions by manipulating whether participants engage in deep or shallow encoding by focusing on the meaning of the action or simply on whether it was performed with one or two hands. While the former should lead to higher memory than the latter (though note that Zimmer and Engelkamp (1999) did not observe substantive increase in memory for enacted actions after deep encoding relative to shallow encoding), it is possible that the encoding focus also affected encoding of source features. The origin of resulting changes in source memory performance can then not be clearly identified.

One approach would be to use the modelling to predict performance in source memory tasks. Berry, Shanks, Li, Rains, and Henson (2010) proposed a single-system model for priming and recognition. Under this model, priming and recognition are modelled with the same underlying uni-dimensional memory distributions, with different levels noise for the priming and recognition task accounting for the seeming dissociations reported in the literature of priming being present even when recognition is absent. Extending such a model for source memory performance (with the core idea of a two-dimensional or one-dimensional account identical), may allow clear predictions of possible dissociations of item and source memory to test if source memory can be easily accounted for by such a uni-dimensional model.

Appendix A

Additional materials for Chapter 2

A.1 Examples of actions

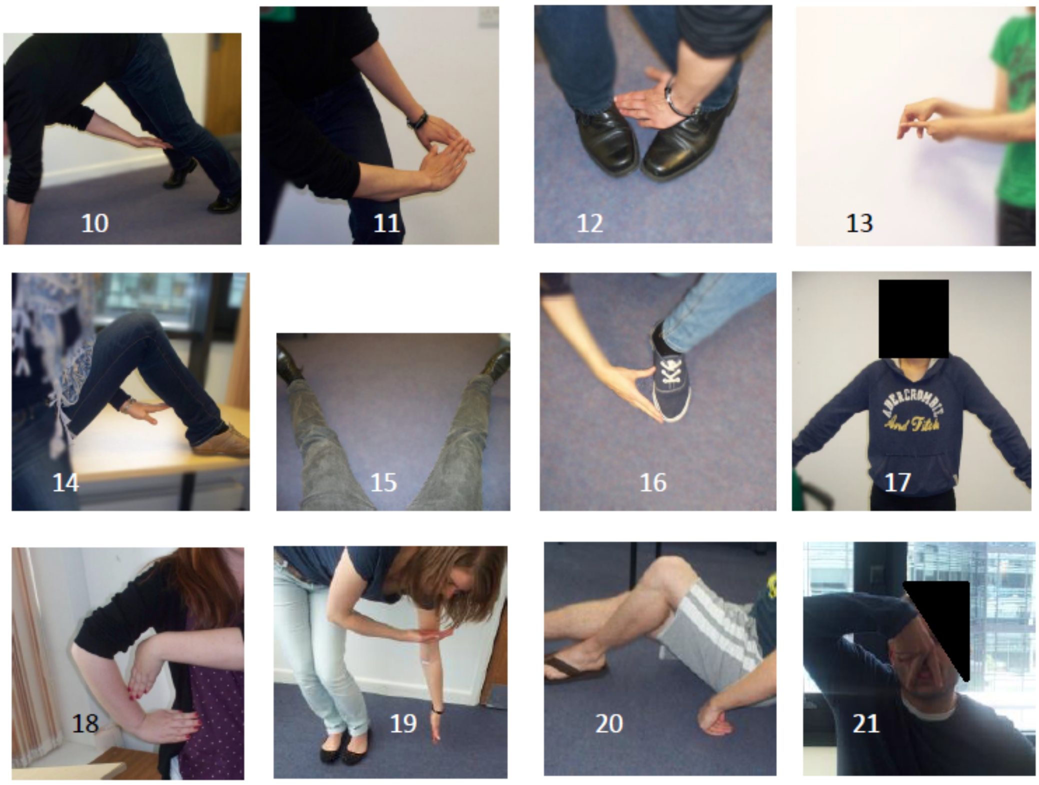

Figure A.1. Some example actions participants produced in response to ‘A’ in Experiments 1 through 3

Appendix B

Additional materials for Chapter 3

B.1 Sets of objects

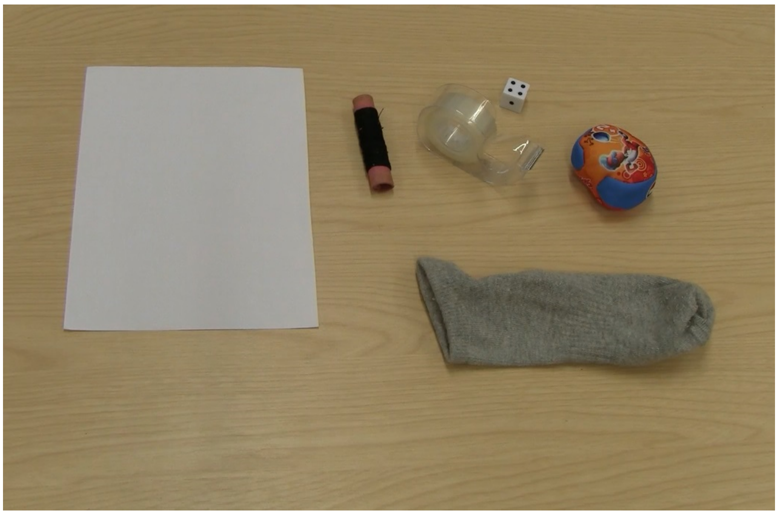

Figure B.1. Objects used in Experiment 4, 5, 6 and 7. Set 1: Paper, Tape dispenser, Dice, Spool of thread, Ball, Sock

Figure B.2. Objects used in Experiment 4, 5, 6 and 7. Set 2: Mug, Ribbon, Card, Vase, Pocket knife, Elastic band

Figure B.3. Objects used in Experiment 4, 5, 6 and 7. Set 3: Box, Car, Glove, Headphones, Post-its, Hook

Figure B.4. Objects used in Experiment 4, 5, 6 and 7. Set 4: Bottle, Cloth, Tub, Peg, Eraser, Pen

Figure B.5. Objects used in Experiment 4, 5, 6 and 7. Set 5: Frisbee, Straw, Stapler, Sharpener, Ring, Tissues

Figure B.6. Objects used in Experiment 4, 5, 6 and 7. Set 6: Book, Cup, Highlighter, Band, Screw, Chain

B.2 Model-fitting

B.2.1 Two-High-Threshold model

I used the approach shown by Bayen et al. (1996) to estimate the best fitting model for participants' aggregate data in Experiment 4 (also see Bayen et al., 2000; Schütz & Bröder, 2011). The aggregate data is shown in Table B.1. I also report the fit for the sum of participants' individual data.

I defined restrictions for the parameters (shown in Figure 3.2 in the main body of the thesis) for the submodels I wanted to test based on theoretical expectations. I fitted the model separately for the Motor and Visual group in line with the model-free analysis, to avoid the difficulty of theoretical inferences based on different behaviour in the control conditions. Rather than also fitting the model separately for Control and Interference conditions, I decided to include the impact of interference on retrieval directly in the model (following Schütz & Bröder, 2011) by restricting the parameter space for the item memory parameter while allowing it to vary for the source memory and guessing parameters. After finding the best-fitting submodel given those theoretical expectations, I will test whether the predictions hold and interference conditions need to be fitted by separate parameters. This process will identify the best-fitting model to compare against 2D-SDT and 1D-SDT models.

The most restricted model under the theoretical assumptions of an influence of interference on source memory and guessing parameters is one defined by 7 parameters ($D_{C,P} = D_{C,O} = D_{C,N} = D_{I,P} = D_{I,O} = D_{I,N}$, $d_{C,P} = d_{C,O}$, $d_{I,P} = d_{I,O}$, $a_C = g_C$, $a_I = g_I$, b_C,b_I), with C indicating the control and I indicating the interference condition. In this model, the same item parameters are assumed for all sources of items and control and interference condition. Source memory is assumed to be equal for performed and observed actions but vary with interference. Source guessing parameters are set equal for detected and undetected items. But they, and the item guessing parameter, are allows to vary.

The number of possible identifiable models rises sharply given the additional degrees of freedom when modelling control and interference condition at the same time. The parameter restrictions of all the submodels I am testing at this stage are shown in Table B.2. The submodels are based on models suggested in Bayen et al. (1996). The submodels I am testing in this first stage are not an exhaustive list of possible submodels. Rather, they include theoretically-based parameter restrictions. For example, all tested submod-

Table B.1. Aggregate frequency of responses in Experiment 4

	Visual group			Motor group		
Responses	Performed	Observed	Novel	Performed	Observed	Novel
Control						
‘Own’	250	71	56	223	88	81
‘Partner’	89	217	119	95	199	104
‘New’	65	116	237	93	124	229
Interference						
‘Own’	257	60	63	252	78	63
‘Partner’	90	232	115	94	206	100
‘New’	60	115	234	67	129	251

Note. Based on 23 participants in each experimental group and on average 18 items in each for each item type.

els allow for separate guessing parameters for control and interference conditions. For example, I am assuming $a = g$ in all models, that is source guessing is equal for item detected as old and items not detected as old (following Schütz & Bröder, 2011).

I fit the models using the *MPTinR* package (Singmann & Kellen, 2013) to the aggregate data. Table B.3 shows the models I tested with the goodness-of-fit statistic G^2 indicating how well the models each account for the data. G^2 is defined as $G^2 = 2 * ln(L)$, where L denotes the maximized value of the likelihood function, and is asymptotically χ^2-distributed. The degrees of freedom to test the fit of the model against the observed data are defined as the difference of the numbers of parameters in the model and the number of observations that are fitted. If the data significantly deviates from the model (G^2 is significant), the model fit is not adequate.

For both the visual and motor group, submodel 2htm.4b provides the best fit to the data. While this fit is adequate in the Visual group, it is marginal in the Motor group. The gains in fit compared to more restricted models such as model 2htm.2b are very small. In a second step, I will now test all parameter restrictions of the winning submodel 2htm.4b directly to identify if existing restrictions are warranted, i.e. if fit improves with a less restricted model or if a more restricted model (such as model 2htm.2b) can account equally well for the data. The aim of this step is to identify the significantly best fitting model that can be defined by the least parameters, on the basis of fit alone.

Table B.2. 2HTM submodels with parameter restrictions

Model	Parameters	D	d	a, g	b
2htm.1	7	$D_{C,P} = D_{C,O} = D_{C,N} = D_{I,P} = D_{I,O} = D_{I,N}$	$d_{C,P} = d_{C,O}$; $d_{I,P} = d_{I,O}$	$a_C = g_C$; $a_I = g_I$	b_C ; b_I
2htm.2a	8	$D_{C,O} = D_{I,O}$; $D_{C,P} = D_{C,N} = D_{I,P} = D_{I,N}$	$d_{C,P} = d_{C,O}$; $d_{I,P} = d_{I,O}$	$a_C = g_C$; $a_I = g_I$	b_C ; b_I
2htm.2b	8	$D_{C,P} = D_{I,P}$; $D_{C,O} = D_{C,N} = D_{I,O} = D_{I,N}$	$d_{C,P} = d_{C,O}$; $d_{I,P} = d_{I,O}$	$a_C = g_C$; $a_I = g_I$	b_C ; b_I
2htm.3	9	$D_{C,P} = D_{C,O} = D_{C,N} = D_{I,P} = D_{I,O} = D_{I,N}$	$d_{C,P}$; $d_{I,P}$; $d_{C,O}$; $d_{I,O}$	$a_C = g_C$; $a_I = g_I$	b_C ; b_I
2htm.4a	10	$D_{C,O} = D_{I,O}$; $D_{C,P} = D_{C,N} = D_{I,P} = D_{I,N}$	$d_{C,P}$; $d_{I,P}$; $d_{C,O}$; $d_{I,O}$	$a_C = g_C$; $a_I = g_I$	b_C ; b_I
2htm.4b	10	$D_{C,P} = D_{I,P}$; $D_{C,O} = D_{C,N} = D_{I,O} = D_{I,N}$	$d_{C,P}$; $d_{I,P}$; $d_{C,O}$; $d_{I,O}$	$a_C = g_C$; $a_I = g_I$	b_C ; b_I

D - item parameter
d - source parameter
b - item guessing parameter
a and g - source guessing parameters
Subscripts: P - Performed actions, O - Observed actions, C - Control condition, I - Interference condition

Table B.3. Model fits of 2HT models in Experiment 4

Model	Parameters	Sum of individual fits		Aggregate fit	
		G^2	p	G^2	p
Visual group					
2htm.1	7	275.38	<.001	59.99	<.001
2htm.2a	8	226.24	<.001	24.46	<.001
2htm.2b	8	149.92	<.001	1.74	.78
2htm.3	9	253.66	<.001	56.33	<.001
2htm.4a	10	218.57	<.001	24.56	<.001
2htm.4b	10	132.18	<.001	0.19	.91
Motor group					
2htm.1	7	229.10	<.001	43.57	<.001
2htm.2a	8	196.20	<.001	15.77	.003
2htm.2b	8	138.95	.001	8.10	.09
2htm.3	9	201.46	<.001	40.26	<.001
2htm.4a	10	180.74	<.001	15.70	<.001
2htm.4b	10	111.66	<.001	7.93	.02

Note. Adequate model fits are indicated by bold values in the G^2 columns.

The adequacy of parameter restrictions can be tested by conducting likelihood ratio tests between the model that allows the tested parameter to vary and one that does not, i.e. by comparing nested models via comparison of their negative log-likelihoods (e.g. Maddox & Ashby, 1993). The resulting G^2 value is asymptotically χ^2-distributed with degrees of freedom resulting from the difference in restrictions between the models being compared.

Before I looked at the effect of my interference manipulation at retrieval, I first looked at item memory parameters. Item memory for performed actions was higher than for observed actions (and novel action phrases) in the Visual group, $G^2(1) = 28.07, p < .001$, and in the Motor group, $G^2(1) = 16.17, p < .001$. I made the assumption that the item parameter should not vary between interference conditions. Indeed, allowing item memory parameters to vary between control and interference condition for performed actions, $G^2 = .07, p = .79$, or observed actions, $G^2 < .001, p = .99$, did not significantly increase fit in the Visual group. In the Motor group, the difference between interference conditions was marginally significant for performed actions, $G^2(1) = 3.56, p = .059$, but not for observed actions, $G^2(1) = .52, p = .47$. This is in line with the model-free analysis showing a marginal effect of better item detection under motor interference.

I next looked at source memory parameters. In the winning model 2htm.4b, I did not restrict source memory parameters by source or interference condition. While this model provided a better fit than the more restricted 2htm.2b model which equated the source memory parameter across sources, this gain in fit was not significant in the Visual group, $G^2(2) = .78, p = .68$, or the Motor group, $G^2(2) = .09, p = .96$. At the same time, there was no gain in fit from estimating source memory parameters separately for the control and interference condition in the Visual group, $G^2(2) = .92, p = .63$, or Motor group, $G^2(2) = .79, p = .67$. This means that interference has no effect on participants' source memory.

I next looked at the item guessing parameter. Did participants display a bias to respond 'Old' in either group and did this bias differ between interference conditions? In the Visual group, participants showed a bias to respond 'Old' in both the Control, $G^2(1) = 8.75, p = .003$, and Interference condition, $G^2(1) = 11.43, p < .001$. There was no evidence for an influence of visual interference at retrieval on this bias, $G^2(1) = .10, p = .75$. In the Motor group, participants showed a bias to respond 'Old' in the Control condition, $G^2(1) = 6.74, p = .009$, and in the Interference condition, $G^2(1) = 5.98, p = .015$. There was no evidence for an influence of motor interference at retrieval on this bias, $G^2(1) < .001, p = .98$. Again, model fit is not significantly increased by fitting control and interference conditions separately for item guessing.

Finally, I looked at the source guessing parameters. In the Visual group, source guessing parameters a and g were significantly below 0.5 both in the Control, $G^2(1) = 13.08, p < .001$, and Interference condition, $G^2(1) = 13.11, p < .001$. This means participants displayed a significant bias to guess 'Partner' over 'Own' in the absence of source memory, the 'It had to be you' bias. There was no evidence that visual interference at retrieval affected this bias, $G^2(1) = .20, p = .65$. In the Motor group, there was no evidence for Interference affecting the source guessing parameters, $G^2(1) = .47, p = .49$, and participants overall showed a bias towards guessing 'Partner' in the absence of source identification, $G^2(2) = 6.11, p = .047$ (though, when estimated separately, the bias was not significant in the Control condition, $G^2(1) = 1.43, p = .23$).

The tests of the parameter restrictions reveal that while submodel 2htm.4b performed best of the submodels tested initially, a more parsimonious model that does not fit significantly worse can be defined. That model is defined as follows: (1) the same item memory parameter for observed and novel actions, with a separate item memory parameter for performed actions, (2) the same source memory parameter for performed and observed actions, (3) parameters set equal across control and interference conditions. Th The parameter estimations for that, most restricted, best-fitting 2HTM model are in Table 3.6 in the main body of the thesis. That is the model I will test against the winning 2D-SDT and 1D-SDT model in the final section of Chapter 3.

B.2.2 Two-dimensional signal detection model

I fitted the two-dimensional signal detection model to the aggregate data shown in Table B.1. I used the same overall procedure of first testing a subset of possible models and then specifically testing the parameter restrictions of the winnng model with likelihood ratio tests. In the two-dimensional signal detection model in Figure 3.3 each source of items (X_{Perf} for performed actions, X_{Obsv} for observed actions and X_{Nov} for novel actions) is represented by a bivariate normal distribution. Each distribution is characterized by five parameters as shown in Equations B.1 through B.3.

$$X_{Perf} \sim N(\mu_{Perf}, \Sigma_{Perf}) \quad \text{with} \quad \mu_{Perf} = \begin{bmatrix} \mu_{I,Perf} \\ \mu_{S,Perf} \end{bmatrix} \quad \text{and} \quad \Sigma_{Perf} = \begin{bmatrix} \sigma_I^2 & r_{I,S} \\ r_{I,S} & \sigma_S^2 \end{bmatrix} \tag{B.1}$$

$$X_{Obsv} \sim N(\mu_{Obsv}, \Sigma_{Obsv}) \quad \text{with} \quad \mu_{Perf} = \begin{bmatrix} \mu_{I,Obsv} \\ \mu_{S,Obsv} \end{bmatrix} \quad \text{and} \quad \Sigma_{Obsv} = \begin{bmatrix} \sigma_I^2 & r_{I,S} \\ r_{I,S} & \sigma_S^2 \end{bmatrix} \tag{B.2}$$

$$X_{Nov} \sim N(\mu_{Nov}, \Sigma_{Nov}) \quad \text{with} \quad \mu_{Nov} = \begin{bmatrix} \mu_{I,Nov} \\ \mu_{S,Nov} \end{bmatrix} \quad \text{and} \quad \Sigma_{Nov} = \begin{bmatrix} \sigma_I^2 & r_{I,S} \\ r_{I,S} & \sigma_S^2 \end{bmatrix} \tag{B.3}$$

where μ indicates the mean of the distribution (with μ_I indicating the mean of the distribution on the item memory and μ_S indicating the mean of the distribution on the source memory dimension). Σ indicates variance-covariance matrix of each bivariate distribution which in turn is characterized by the distributions' standard deviation on the the item memory dimension (σ_I) and source memory dimension (σ_S), as well as the correlation between item memory and source memory (r).

Aside from the distributions, the 2D-SDT model also defines decision criteria on the item memory and source memory dimension. In the case of binary decisions on both dimensions, a single criterion orthogonal to the dimension is assumed on each dimension (c_{old} for Old/New decisions and c_{own} for Performed/Observed decisions).

With the combination of 15 parameters defining the distributions and 2 decision criteria, the full 2D-SDT model is defined by 17 parameters in total. I restricted this unequal variance model to its equal variance case with all distributions set to a variance of 1. I additionally assumed the dimensions to be uncorrelated. This means the covariance matrix in the equations above was replaced by the identity matrix for all distributions. Finally, arguably novel items do not contain item or source memory strength. Therefore, the distributions of novel items on both dimensions are defined as standard normal distribution with means of 0. The strength of the distributions for performed and observed items are now defined in relation to novel items.

With those restrictions, the distributions for the three types of item in the present experiment are given in equations B.4 through B.5. This restricted model is defined by a total of 6 parameters (4 for the distributions, 2 criteria). To avoid confusion, I will re-use parameter names D_P, D_O and D_N to indicate item memory strength of performed, observed and novel actions on the item memory dimension, though note that it indicates the mean of a distribution here rather than a probability as in the 2HTM. I will re-use d_P, d_O and d_N to indicate source memory strength of performed, observed and novel actions on the source memory dimension (note again, this indicates the mean of a distribution here).

$$X_{Performed} \sim N(\mu_P, \Sigma_P) \text{ with } \mu_P = \begin{bmatrix} D_P \\ d_P \end{bmatrix} \text{ and } \Sigma_P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{B.4}$$

where:

$$\begin{aligned} D_P &= \mu_{Item,P} - \mu_{Item,N} \\ d_P &= \mu_{Source,P} - \mu_{Source,N} \end{aligned}$$

$$X_{Observed} \sim N(\mu_O, \Sigma_O) \text{ with } \mu_O = \begin{bmatrix} D_O \\ d_O \end{bmatrix} \text{ and } \Sigma_O = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{B.5}$$

where:

$$\begin{aligned} D_O &= \mu_{Item,O} - \mu_{Item,N} \\ d_O &= \mu_{Source,O} - \mu_{Source,N} \end{aligned}$$

$$X_{Novel} \sim N(\mu_N, \Sigma_N) \text{ with } \mu_N = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } \Sigma_N = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{B.6}$$

The responses in the 3AFC source recognition task can be represented by probabilities in the 2D-SDT model. The probabilities for each possible response in the 2D-SDT model are defined by the equations B.7 through B.15.

$$P(\text{'Own'|Performed}) = \Phi_2(a, b | \mu_P, \Sigma_P) \text{ with } a = \begin{bmatrix} c_{old} \\ c_{own} \end{bmatrix} \text{ and } b = \begin{bmatrix} \infty \\ \infty \end{bmatrix} \tag{B.7}$$

Table B.4. Tested 2D-SDT submodels with parameter restrictions

Model	Parameters	D	d	c_{old}	c_{own}
2D.1	7	$D_{C,P} = D_{I,P}$; $D_{C,O} = D_{I,O}$	$d_{C,P} = d_{I,P} = d_{C,O} = d_{I,O}$	$c_{old,C}$; $c_{old,I}$	$c_{own,C}$; $c_{own,I}$
2D.2a	8	$D_{C,P} = D_{I,P}$; $D_{C,O} = D_{I,O}$	$d_{C,P} = d_{I,P}$; $d_{C,O} = d_{I,O}$	$c_{old,C}$; $c_{old,I}$	$c_{own,C}$; $c_{own,I}$
2D.2b	8	$D_{C,P} = D_{I,P}$; $D_{C,O} = D_{I,O}$	$d_{C,P} = d_{C,O}$; $d_{I,P} = d_{I,O}$	$c_{old,C}$; $c_{old,I}$	$c_{own,C}$; $c_{own,I}$
2D.3a	9	$D_{C,P} = D_{I,P}$; $D_{C,O} = D_{I,O}$	$d_{C,P}$; $d_{I,P}$; $d_{C,O} = d_{I,O}$	$c_{old,C}$; $c_{old,I}$	$c_{own,C}$; $c_{own,I}$
2D.3b	9	$D_{C,P} = D_{I,P}$; $D_{C,O} = D_{I,O}$	$d_{C,P} = d_{I,P}$; $d_{C,O}$; $d_{I,O}$	$c_{old,C}$; $c_{old,I}$	$c_{own,C}$; $c_{own,I}$
2D.4	10	$D_{C,P} = D_{I,P}$; $D_{C,O} = D_{I,O}$	$d_{C,P}$; $d_{I,P}$; $d_{C,O}$; $d_{I,O}$	$c_{old,C}$; $c_{old,I}$	$c_{own,C}$; $c_{own,I}$

D - item memory strength of the distributions relative to novel items
d - source memory strength of the distribution relative to novel items
Subscripts: P - Performed actions, O - Observed actions, C - Control condition, I - Interference condition.
c_{old} - decision criterion on the item memory dimension
c_{own} - decision criterion on the source memory dimension

$$P(\text{‘Own’|Observed}) = \Phi_2(a,b|\mu_O,\Sigma_O) \text{ with } a = \begin{bmatrix} c_{old} \\ c_{own} \end{bmatrix} \text{ and } b = \begin{bmatrix} \infty \\ \infty \end{bmatrix} \tag{B.8}$$

$$P(\text{‘Own’|Novel}) = \Phi_2(a,b|\mu_N,\Sigma_N) \text{ with } a = \begin{bmatrix} c_{old} \\ c_{own} \end{bmatrix} \text{ and } b = \begin{bmatrix} \infty \\ \infty \end{bmatrix} \tag{B.9}$$

$$P(\text{‘Partner’|Performed}) = \Phi_2(c,e|\mu_P,\Sigma_P) \text{ with } c = \begin{bmatrix} c_{old} \\ -\infty \end{bmatrix} \text{ and } e = \begin{bmatrix} \infty \\ c_{own} \end{bmatrix} \tag{B.10}$$

$$P(\text{‘Partner’|Observed}) = \Phi_2(c,e|\mu_O,\Sigma_O) \text{ with } c = \begin{bmatrix} c_{old} \\ -\infty \end{bmatrix} \text{ and } e = \begin{bmatrix} \infty \\ c_{own} \end{bmatrix} \tag{B.11}$$

$$P(\text{‘Novel’|Observed}) = \Phi_2(c,e|\mu_N,\Sigma_N) \text{ with } c = \begin{bmatrix} c_{old} \\ -\infty \end{bmatrix} \text{ and } e = \begin{bmatrix} \infty \\ c_{own} \end{bmatrix} \tag{B.12}$$

$$P(\text{‘New’|Performed}) = \Phi_2(f,g|\mu_P,\Sigma_P) \text{ with } f = \begin{bmatrix} -\infty \\ -\infty \end{bmatrix} \text{ and } g = \begin{bmatrix} c_{old} \\ \infty \end{bmatrix} \tag{B.13}$$

$$P(\text{‘New’|Observed}) = \Phi_2(f,g|\mu_O,\Sigma_O) \text{ with } f = \begin{bmatrix} -\infty \\ -\infty \end{bmatrix} \text{ and } g = \begin{bmatrix} c_{old} \\ \infty \end{bmatrix} \tag{B.14}$$

$$P(\text{‘New’|Novel}) = \Phi_2(f,g|\mu_N,\Sigma_N) \text{ with } f = \begin{bmatrix} -\infty \\ -\infty \end{bmatrix} \text{ and } g = \begin{bmatrix} c_{old} \\ \infty \end{bmatrix} \tag{B.15}$$

Here Φ_2 denotes bivariate distributions; a, c and f denote the lower bounds of integration and b, e and g the upper bounds of integration.

The data is defined by 6 degrees of freedom, with the model defined by 6 parameters. As in the multinomial case, I estimated the parameters for the interference conditions (Control, Interference) at the same time, resulting in 12 degrees of freedom and 12 parameters to be estimated. To avoid overfitting of the data, I further restricted the parameter space using the same theoretical reasoning as in the 2HTM case above. I tested the effect of interference on the average source memory strength of performed and observed actions by testing some models that assumed separate and some that assumed joint parameters. All submodels I tested assumed that visual and motor interference would have differential impact on the decision criteria. Beyond that, the tested submodels vary in their assumptions of the relative strength of performed and observed actions on the item memory and source memory dimension. The tested submodels are in Table B.4.

As with the 2HTM above, the 2D-SDT model was fit using the *MPTinR* package (Singmann & Kellen, 2013) by using the equations and restrictions defined above, with the multivariate normal distributions implemented with the *mvtnorm* package (Genz et al., 2016). The models were fit separately to the Visual group and Motor group. Table B.5 shows the model fits of the tested submodels to the data. A non-significant G^2 value indicates a good model fit by showing that data and model do not differ significantly from one another. None of the tested submodels achieved a good model fit.

Several models in both experimental groups provide equally good fit, already suggesting that the parameter restrictions specified in the submodels do not significantly affect model fit. For ease of the description I will test the parameter restrictions of model 2D.2a explicitly. In the Visual group, it is the least complex of the models providing equal fit to the data. In the Motor group, 2D.3a provides a slightly better fit. I will test whether the reduction in fit compared to that more complex model is significant in the parameter restriction tests below.

I tested the parameter restrictions of the model by comparing the model fit to models that do not have the same restriction. On the item memory dimension, there was no evidence that interference affected the strength of performed actions in the Visual group, $G^2(1) < .001, p > .99$, or in the Motor group, $G^2(1) = .89, p = .35$. Similarly, there was no evidence that interference affected the strength of observed actions in the Visual group, $G^2(1) = .01, p = .91$, or in the Motor group, $G^2(1) = .11, p = .74$. The c_{old} criterion was not significantly changed by interference in the Visual group, $G^2(1) = .12, p = .73$, or in the Motor group, $G^2(1) = .004, p = .95$.

Table B.5. Model fits of 2D SDT models in Experiment 4

Model	Parameters	Sum of individual fits G^2	Sum of individual fits p	Aggregate fit G^2	Aggregate fit p
Visual group					
2D.1	7	646.24	<.001	358.06	<.001
2D.2a	8	345.89	<.001	38.80	<.001
2D.2b	8	638.91	<.001	357.32	<.001
2D.3a	9	338.13	<.001	38.80	<.001
2D.3b	9	335.13	<.001	38.80	<.001
2D.4	10	330.52	<.001	38.80	<.001
Motor group					
2D.1	7	528.56	<.001	239.77	<.001
2D.2a	8	292.66	<.001	25.28	<.001
2D.2b	8	517.58	.001	238.78	<.001
2D.3a	9	281.56	<.001	24.00	<.001
2D.3b	9	284.75	<.001	25.28	<.001
2D.4	10	276.00	<.001	24.00	<.001

Note. Adequate model fits are indicated by bold values in the G^2 columns.

On the source memory dimension, the strength of observed actions did not exceed 0 in either experimental group. This means that under this model, observed actions did not contain source-specifying information. This was not affected by interference in either experimental group. Interference did not affect the source memory strength of performed actions in either experimental group either, all $G^2 < .001, p > .99$. Neither did the the decision criterion change across interference conditions in the Visual group, $G^2(1) = .03, p = .86$, or in the Motor group, $G^2(1) = .16, p = .69$.

This suggests that a model that does not assume an effect of interference on any of the parameters provides as good a fit as a model that assumes an effect of interference, and the more parsimonious model will be preferred. The most parsimonious model that does not provide a significantly worse fit than more complex models is therefore defined by the following parameter restrictions: (1) separate estimation of the mean item memory strength of performed and observed actions, with novel actions set to a mean strength of 0, (2) separate estimation of the mean source memory strength of performed and observed actions, with novel actions set to a mean strength of 0, and (3) estimating the same parameters for control and interference condition. The parameter estimations for the most parsimonious model is shown in Table 3.7 in the main body of the thesis. That model will be the model I will test against the best-fitting 2HTM and 1D-SDT model for best fit overall.

B.2.3 One-dimensional signal detection model

I fitted the one-dimensional signal detection model (1D-SDT) to the aggregate data shown in Table B.1. In the one-dimensional signal detection model in Figure 3.4 each source of items (X_{Perf} for performed actions, X_{Obsv} for observed actions and X_{Nov} for novel actions) is represented by an univariate normal distribution. Each distribution is characterized by two parameters as shown in Equations B.1 through B.3.

$$X_{Perf} \sim N(\mu_{Perf}, \sigma_{Perf}) \tag{B.16}$$

$$X_{Obsv} \sim N(\mu_{Obsv}, \sigma_{Obsv}) \tag{B.17}$$

$$X_{Nov} \sim N(\mu_{Nov}, \sigma_{Nov}) \tag{B.18}$$

Given those equations, the probabilities of responses to items can be estimated by the following equations B.19 through B.27.

$$P(\text{‘Own’|Performed}) = \Phi(\frac{\mu_P - c_{own}}{\sigma_P}) \tag{B.19}$$

$$P(\text{‘Own’|Observed}) = \Phi(\frac{\mu_O - c_{own}}{\sigma_O}) \tag{B.20}$$

$$P(\text{‘Own’|Novel}) = \Phi(-c_{own}) \tag{B.21}$$

$$P(\text{‘Partner’|Performed}) = \Phi(\frac{c_{own} - \mu_P}{\sigma_P}) - \Phi(\frac{c_{old} - \mu_P}{\sigma_P}) \tag{B.22}$$

$$P(\text{‘Partner’|Observed}) = \Phi(\frac{c_{own} - \mu_{Obsv}}{\sigma_O}) - \Phi(\frac{c_{old} - \mu_O}{\sigma_O}) \tag{B.23}$$

$$P(\text{‘Partner’|Novel}) = \Phi(c_{own}) - \Phi(c_{old}) \tag{B.24}$$

Table B.6. 1D SDT submodels with parameter restrictions

Model	Parameters	D	c_{old}	c_{own}
1D.ev.1	5	$D_{C,P} = D_{I,P} = D_{C,O} = D_{I,O}$	$c_{old,C}$; $c_{old,I}$	$c_{own,C}$; $c_{own,I}$
1D.ev.2	6	$D_{C,P} = D_{I,P}$; $D_{C,O} = D_{I,O}$	$c_{old,C}$; $c_{old,I}$	$c_{own,C}$; $c_{own,I}$
1D.ev.3a	7	$D_{C,P} = D_{I,P}$; $D_{C,O}$; $D_{I,O}$	$c_{old,C}$; $c_{old,I}$	$c_{own,C}$; $c_{own,I}$
1D.ev.3b	7	$D_{C,P}$;$D_{I,P}$; $D_{C,O} = D_{I,O}$	$c_{old,C}$; $c_{old,I}$	$c_{own,C}$; $c_{own,I}$
1D.ev.4	8	$D_{C,P}$; $D_{I,P}$; $D_{C,O}$;$D_{I,O}$	$c_{old,C}$; $c_{old,I}$	$c_{own,C}$; $c_{own,I}$

D - average memory strength of the distributions relative to novel items
c_{old} - decision criterion for the Old/New decision
c_{own} - decision criterion for the Performed/Observed decision
Subscripts: P - Performed actions, O - Observed actions, C - Control condition, I - Interference condition.

$$P(\text{'New'|Performed}) = \Phi(\frac{c_{old} - \mu_P}{\sigma_P}) \tag{B.25}$$

$$P(\text{'New'|Observed}) = \Phi(\frac{c_{old} - \mu_O}{\sigma_O}) \tag{B.26}$$

$$P(\text{'New'|Novel}) = \Phi(c_{old}) \tag{B.27}$$

Here Φ designates a univariate normal distribution with mean μ and standard deviation σ. The subscripts designate performed (P), observed (O) and novel (N) items. c_{old} is the Old/New decision criterion and c_{own} is the Performed/Observed decision criterion.

As with the 2HTM and 2D-SDT model, I restricted the parameter space in accordance with theoretical predictions of the experiment. The restrictions for the submodels are set out in Table B.6. As with the 2D-SDT model, I initially tested the equal variance version of the model. Some models assumed that the average memory strength of the distributions could vary with visual or motor interference. Parameter estimates for the decision criteria were made separately for control and interference condition for all models. In a first step, I will select the best-fitting submodel out of the models I tested for further tests of the parameter restrictions.

The model fits for the models in Table B.6 are given in Table B.7. As was the case with 2D-SDT models, none of the models achieve adequate fit. Aside from the model setting the memory strength equal across performed and observed actions, all models performed similarly well.

Submodel 1D.4 provided the best fit for the aggregate data in both the Visual and Motor group. This was the least restricted submodel I tested. I will now looked at whether the

Table B.7. Model fits of 1D SDT models in Experiment 4

		Sum of individual fits		Aggregate fit	
Model	Parameters	G^2	p	G^2	p
Visual group					
1D.1	5	667.13	<.001	397.40	<.001
1D.2	6	355.23	<.001	141.39	<.001
1D.3a	7	333.61	<.001	140.78	<.001
1D.3b	7	314.03	<.001	141.10	<.001
1D.4	8	283.23	<.001	140.75	<.001
Motor group					
1D.1	5	550.75	<.001	280.55	<.001
1D.2	6	322.96	<.001	120.29	<.001
1D.3a	7	304.90	<.001	119.66	<.001
1D.3b	7	281.44	<.001	109.90	<.001
1D.4	8	263.04	<.001	108.84	<.001

Note. Adequate model fits are indicated by bold values in the G^2 columns.

increase in fit relative to more restricted models is significant.

In submodel 1D.4, I estimated a different mean memory strength for performed and observed actions and for both interference conditions. In fact, submodel 1D.4a fits the data significantly better than the model assuming the same mean memory strength for both distributions in both control and interference condition in the Visual group, $G^2(3) = 128.33, p < .001$, and the Motor group, $G^2(3) = 85.85, p < .001$.

In the Visual group, there is no evidence for visual interference affecting the mean memory strength of performed, $G^2(1) = .17, p = .68$, and observed actions, $G^2(1) = .01, p = .91$. In fact, the model estimating the same parameter for control and interference conditions for both sources of items provides no significantly worse fit than submodel 1D.4, $G^2(2) = .32, p = .85$.

In the Motor group, the separate estimation of memory strength for observed actions in the control and interference condition does not significantly improve fit over a more restricted model, $G^2(1) = .53, p = .47$. This more restricted model (submodel 1D.3b) therefore provides a similarly good fit while requiring fewer parameters. In fact, this more restricted submodel provides a better fit than the model that assumes that interference foes not affect the mean memory strength of either performed and observed actions (submodel 1D.2), $G^2(1) = 5.20, p = .02$.

The c_{old} criterion does not significantly change when actions are retrieved under interference in the Visual group, $G^2(1) = .16, p = .67$, or in the Motor group, $G^2(1) = 1.25, p = .26$. Similarly, the c_{own} criterion does not significantly change with the interference condition in the Visual group, $G^2(1) = .02, p = .90$, or the Motor group, $G^2(1) = 2.44, p = .12$.

Table B.8. Comparisons of model fits of the 2HTM, 2D-SDT and 1D-SDT models in Experiment 4

		Individual data					Aggregate data		
Model	Parameters	G^2	p	best describes		AIC	G^2	p	AIC
Visual group									
2HTM	5	213.47	.003	23		443.47	2.21	.95	12.21
2D-SDT	6	364.94	<.001	0	5	640.94	39.10	<.001	51.10
1D-SDT	5	295.50	<.001		18	525.50	74.86	<.001	84.86
Motor group									
2HTM	5	190.11	.06	22		420.11	10.18	.18	20.18
2D-SDT	6	308.19	<.001	1	4	584.19	25.60	<.001	37.60
1D-SDT	5	260.54	<.001		19	490.54	79.15	<.001	89.15

G^2 - goodness of fit
AIC - Akaike Information Criterion

Finally, I looked at whether the assumption of the equal-variance model provides sufficient fit or if the model fit can be improved by assumptions of unequal-variance. Assuming a standard deviation of 1 for observed actions but allowing parameter estimation of the standard deviation of performed actions increased the fit significantly in the Visual group, $G^2(1) = 33.38, p < .001$, and the Motor group, $G^2(1) = 20.33, p < .001$. Parameter estimation of the standard deviation of observed actions (with standard deviation of performed actions held constant at 1) did not significantly improve fit in the Visual group, $G^2(1) < .001, p > .999$, or the Motor group, $G^2(1) = .02, p = .88$. While I reported above, that estimating the parameters for the mean of the performed actions distribution separately for control and motor interference condition improves the fit of the model, this is not the case when both tested models are unequal-variance models. In that case, the model assuming no effect of interference on performed and observed actions, but assuming a broader standard deviation for performed actions, provides no worse fit than the model fitting separate parameters for the interference conditions for performed actions, $G^2(1) = 2.77, p = .10$.

To summarize the results, the model that provides the best fit to the data differ for both the Visual and the Motor group is characterized by (1) separate estimation of memory strength for performed and observed actions, with parameters set equal across control and interference condition, (2) a standard deviation of 1 for observed actions with the standard deviation of performed actions free to vary, (3) no effect of interference on any parameters in the model. The parameter estimates for this best-fitting 1D-SDT model are in Table 3.8 in the main body of the thesis. I will test this model against the best-fitting 2HTM and 2D-SDT model in the final section of the model-based analysis.

B.3 Comparison of best-fitting 2HTM, 2D-SDT and 1D-SDT models

Table B.8 shows the formal comparison of all three models by AIC (with Table 3.9 in the main body of the thesis only showing BIC). As the tables shows, both AIC (defined as $AIC = 2 * k - 2 * ln(L)$ where k denotes the number of parameters and L the maximized value of the likelihood function) and BIC come to the same conclusions.

Appendix C

Additional materials for Chapter 4

C.1 Model-fitting

I adapted the one-dimensional signal detection model for the extended selective recall task (also see Hollins et al., 2016; Marsh & Bower, 1993). The underlying equations for the distributions and boundaries in the one-dimensional signal detection model are equivalent to those defined in Chapter 3 under the assumption that even with a change in retrieval task, the underlying memorial distributions and mechanisms of memory should not change. Figure C.1 illustrates the case for the one-dimensional model based on the extended recall data. I fitted every experiment in Chapter 4 twice: (1) I only fitted the final report data, and (2) I fitted the extended recall (availability and monitoring) data.

Adapting signal detection models to recall data

Figure C.1 illustrates the one-dimensional signal detection model. As I explained in the previous chapter, under this model participants' source decisions are based on the overall memory strength of items. If an item's strength exceeds the output criterion c_{out}, the item is identified as 'old', otherwise it is identified as novel. If an item's strength additionally exceeds the $c_{performed}$ criterion, it is identified as a performed action, otherwise it is identified as an observed action. In a source recognition task, those decisions are reflected by participants identifying every item as performed, or observed, or new.

In a standard (non-extended) source-cued recall task, participants provide information about those same decisions by reporting items at test, rather than responding to items presented to them. When participants are asked to recall performed actions, for example, under this one-dimensional signal detection model, they will report actions that exceed the $c_{performed}$ criterion. Some of those actions will be correct, some will be observed actions and some will be novel (mapping onto the tails of the distributions exceeding this criterion). In the Recall observed task, on the other hand, participants report actions that exceed the c_{out} criterion but not the $c_{performed}$ criterion. This way the recall task provides

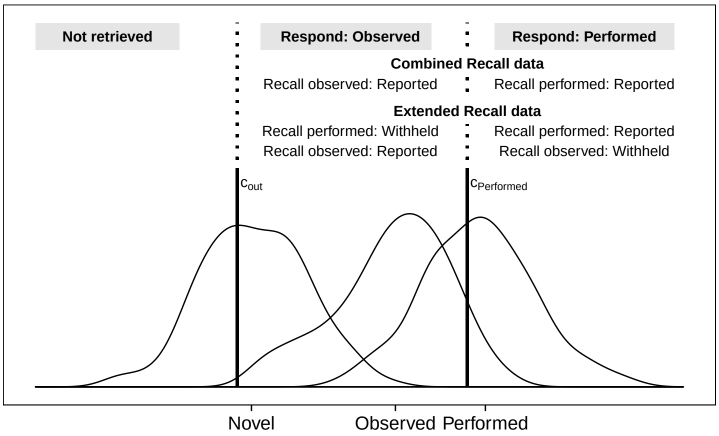

Figure C.1. Illustration of the single-dimension model of source memory in a recall task for novel items and two sources of old items (performed actions and observed actions), with decision criteria are c_{out} and$c_{performed}$. At the top, illustration of the observed data the model is fitted to when fitting only the final report data or the full extended recall data

similar information to the recognition task about the distributions underlying memory performance (though arguably the c_{old} criterion in a recognition task and c_{out} criterion in a recall task are not necessarily equivalent).

The probabilities of a response given the distribution (to account for the areas under each distribution given its boundaries and determine the parameters of the model) are calculated as in Chapter 3 (Equations B.19 to B.27). In a recognition task, the absolute frequencies (and hence probabilities) of any possible response given the item distribution are known since all participants classify all items. In a recall task, crucially, the absolute number of novel actions that could be retrieved is unknown. There is no obvious rationale for estimating the number of novel items available to any single participant or how those numbers may differ between participants given the pattern of responding they have shown (see the discussion of the guessing adjustment in Chapter 2).

To estimate if this model could account for the data produced in the experiments in Chapter 4, I used a different approach than the maximum likelihood estimation in the previous chapter. Instead, I minimized across frequencies of items derived from proportions of responses given a recall task. In the empirical data, participants, for example, only recall performed actions in the Recall performed task (the section to the right of the $c_{performed}$ criterion in Figure C.1). The relative proportion of a correct performed response in the Recall performed task is determined as the number of performed actions over the total of all actions reported in the Recall performed task. The equations for this are given in

Equation C.1 through C.6.[1] The proportions of all responses in a recall task then sum to 1.

Importantly, those proportions indicate the probability of an item given a response, rather than the probability of a response given an item. I therefore similarly calculated the same proportions in the model fitting procedure. Since the number of performed, observed and novel actions that were produced at retrieval at all are known, those proportions can in turn be used to calculate the frequency of items, given the boundaries of the space. I then minimized the difference (residual sum of squares) between observed and predicted frequencies of items.

$$p(\text{Performed}|\text{`Performed'}) = \frac{P(\text{`}P'|P)}{P(\text{`}P'|P) + P(\text{`}P'|O) + P(\text{`}P'|N)} \tag{C.1}$$

$$p(\text{Observed}|\text{`Performed'}) = \frac{P(\text{`}P'|O)}{P(\text{`}P'|P) + P(\text{`}P'|O) + P(\text{`}P'|N)} \tag{C.2}$$

$$p(\text{Novel}|\text{`Performed'}) = \frac{P(\text{`}P'|O)}{P(\text{`}P'|P) + P(\text{`}P'|O) + P(\text{`}P'|N)} \tag{C.3}$$

$$p(\text{Performed}|\text{`Observed'}) = \frac{P(\text{`}O'|P)}{P(\text{`}O'|P) + P(\text{`}O'|O) + P(\text{`}O'|N)} \tag{C.4}$$

$$p(\text{Observed}|\text{`Observed'}) = \frac{P(\text{`}O'|O)}{P(\text{`}O'|P) + P(\text{`}O'|O) + P(\text{`}O'|N)} \tag{C.5}$$

$$p(\text{Novel}|\text{`Observed'}) = \frac{P(\text{`}O'|N)}{P(\text{`}O'|P) + P(\text{`}O'|O) + P(\text{`}O'|N)} \tag{C.6}$$

Rather than fitting the model by maximum likelihood as in the previous chapter, I therefore simply determined which parameter estimates can best account for the recall data, assuming that this particular model underlies performance in the recall task. Assuming an equal-variance model, the parameters I am looking to fit are the mean of the performed-actions distribution (D_P), the mean of the observed-actions distribution (D_O) and the placement of c_{old} and $c_{Performed}$. With the distribution novel items set to 0, as per convention, the strength of the distributions of performed and observed actions will be

---

[1] Note 'Own' responses in Experiment 4 correspond to 'Performed' recall here, indicated by 'P', and 'Partner' responses correspond to 'Observed' recall, indicated by 'O'. Novel actions are indicated by 'N'

calculated relative to that baseline.

I adapted the two-dimensional signal detection model to recall data in the same way as the one-dimensional signal detection model. Given the assumption of two memory strength dimensions (item memory and source memory), the number of parameters that are estimated is higher. Using the same approach of minimizing across frequency (estimates) based on proportions of recalled items given a task, I will estimate parameters for the mean strength of performed and observed actions on the item dimension (D_P and D_O) and on the source dimension (d_P and d_O), as well as the decision criterion on the item dimension (c_{old}) and the decision criterion on the source dimension ($c_{Performed}$).

C.1.1 Data

Fitting the final report data

Participants in a recognition task are asked to respond ‘Performed’, ‘Observed’ or ‘Novel’ to every item presented to them, and the parameters can be easily estimated given that data, as I showed in Chapter 3. On the other hand, participants in a standard (non-extended) source-cued recall task report only items they believe to have originated in one source. In the Recall performed task this means they provide the data for everything beyond the $c_{Performed}$ criterion, and in the Recall observed task they provide the data for everything between the c_{old} and $c_{Performed}$ criterion. In the experiments in Chapter 4, this selective source-cued report is equivalent to the responses participants report as task-compliant, i.e., their final report, excluding items participants chose to withhold as not task-compliant. If the data is combined across retrieval tasks, all proportions specified in Equations C.1 to C.6 would be accounted for. I therefore combined the final report data to estimate the best-fitting parameters for the data, given the assumption that the underlying memorial distributions are of the shape in Figure 4.1. The combination of the data is indicated in the top half of Figure 4.1.

The most critical assumption of this approach is that responses in the Recall performed and Recall observed task do not overlap. In other words, that items participants falsely report as performed are not also items participants correctly report as observed in the Recall observed task. Based on this assumption, the distribution of performed and observed actions can be fully defined by combining correct responses in the Recall performed task, false performed responses in the Recall observed data and the number of performed actions not reported in either task. The same is true for observed actions. This assumption may be premature.

Fitting the extended recall data

The extended recall task, as indicated in Figure C.1, provides the data necessary to account for every slice of the distribution (bar the novel items not produced at recall). Note, participants were not required to specify the source of the actions they withheld from report, beyond indicating that they were not task-compliant. In the one-dimensional signal detection model this means that in the Recall performed task, reported actions account for the items that exceed the $c_{performed}$ criterion, while withheld actions account for the items that exceed the c_{out} criterion but not the $c_{performed}$ criterion. While the latter is qualitatively different to reporting actions as ‘Observed’ since participants may believe those actions to be observed or novel actions, it is mathematically equivalent given the model since items are recalled (exceed c_{out}) but are not reported as performed (do not exceed $c_{performed}$). Each participant therefore provides the information necessary to account for all data points that could be fit with this model.

However, participants were asked to orient towards performed or observed actions by their retrieval task instruction. What follows is that the ‘withheld’ set of items in the Recall observed task is probably not identical in composition to the ‘reported’ set of items in the Recall performed task. Assuming that orienting towards one source would skew production of items at test towards that source, I estimated parameters simultaneously for participants in the Recall performed and the Recall observed task. The mean strength of observed actions is then based on participants’ retrieval of observed actions, both when they were the target of retrieval, and not the target of retrieval. Since participants were asked to only recall performed or only recall observed items, I fitted the model to the aggregate data.

In addition to fitting the final report data, I therefore also fitted the extended recall data to determine parameter estimates that provide the closest fit of the model to the data.

C.1.2 Procedure

Given the limited number of data points provided by the data and the complexity of the parameter space in both models, I restricted the parameters for both one- and two-dimensional models to the most parsimonious and theoretically motivated versions of those models. I fitted the equal-variance one-dimensional model (the unequal variance model did not provide significant gains in model fit). In this model I estimated the best fitting parameters for the means of the performed and observed actions distribution relative to novel items at 0 and the c_{out} and $c_{performed}$ criteria, separately for sets with experimenter-provided actions and sets with participant-generated actions. This resulted in 8 parameters to account for 12 data points to be fit when combining final report data across retrieval tasks and 8 parameters to account for 24 data points to be fit when looking at the extended recall data.

I fitted the equal-variance two-dimensional model with independent dimensions. By estimating means of performed and observed actions relative to novel items at 0 on the item strength and source strength dimension, as well as c_{out} and $c_{performed}$ separately for both types of sets, I would have to estimate 12 data points based on 12 parameters. I restricted the parameter space further by testing the specific diagnostic features model I proposed in the introduction. This source-monitoring model would suggest that generation critically affects the source strength of items, regardless of an effect on the overall memory strength. I therefore set the distributions and c_{out} on the item strength dimension equal for sets with experimenter-provided actions and those with participant-generated actions. Meanwhile, on the source strength dimension, I allowed the means of the distributions and the criterion to vary between conditions. This results in 9 parameters to account for 12 data points for the combined estimation and 9 parameters to account for 24 data points for the extended recall case.

I set the average strength of novel items to 0 for both models, as novel items are not encoded and so do not contain item memory or source memory strength. I did not restrict the parameters for the average strength of the distribution to exceed 0.

Appendix D

Additional materials for Chapter 5

D.1 List of action items

Take out a tissue

Open the safety pin

Stretch the rubber band

Take the CD out of the case

Cut off a piece of thread

Squeeze the sponge

Put the candle in the glass

Push the toy car

Put the toothbrush in the cup

Slide the card into the holder

Tear off a piece of toilet paper

Open the buckle of the belt

Put the glove on your hand

Erase the pencil on the paper

Pull out the ruler

Open the jar

Unbend the paperclip

Shake the bottle

Roll up the newspaper

Unscrew the pen

Put a pin into the pad

Close the sunglasses

Tear off a piece of paper from the pad

Wrap up the ribbon

Ring the bell

Unlock the lock

Operate the calculator

Take a match out of the match box

Take off a sticky note

Take a cup from the stack

Take the lid off the box

Screw off the cap of the deodorant

Roll the dice

Set the alarm clock

Put the USB cable into the plug

Tear the paper

Take off the highlighter’s cap

Sharpen the pencil

Put the pack of noodles into the bag

Shuffle the cards

Open the book

Fold the kitchen towel

Tie a knot in the twine

Put the soap in the dish

Turn the sock inside out

Polish the spoon

Take the bulb out of the box

Put the egg in the eggcup

Hole punch the paper

Put on the watch

Put the teabag in the cup

Tear off a piece of tape

Put the card in the envelope

Open the pencil case

Operate the remote control

Unwrap a candy

Dial a number on the phone

Shake the pins

Put the fork on the plate

Shine the flashlight

D.2 Model-based analysis

I fitted response bias instantiations of the Two-High-Threshold (2HTM), one-dimensional (1D-SDT) and two-dimensional signal detection model (2D-SDT) for the data in Experiment 8 (collapsed across visual perspective, testing only no observation and repeated observation manipulations). As in Chapter 3, I will first identify the best-fitting model of each type by goodness of fit G^2 for the aggregate data. I will then compare the fits of the three best-fitting models against one another.

Two-High-Threshold model

The basic formulation of the model follows the one presented in Chapter 3. The assumption of multinomial tree processing models is that responding arises from certainty and uncertainty states. In certainty states, participants classify items based on correctly identifying the origin of the item. In uncertainty states, participants guess, with responses based on guessing leading to correct or incorrect identifications of items. The graphical representation of the model is in Figure 3.2 on page 116, with read actions replacing observed actions for this experiment.

Participants responded to actions that they had performed in the initial encoding phase, read in the initial encoding phase or that had been not presented in the initial encoding phase. Participants could respond ‘performed’, ‘read’ or ‘not presented’. The manipulation of observation frequency in Experiment 8 is then equivalent to the manipulation of the retrieval interference in Experiment 4. Not observing the experimenter perform actions between encoding and retrieval serves as the control or baseline condition. Observing the experimenter perform an action repeatedly leads to the critical effect this observation condition has on responding that the model has to accommodate.

I will use the same terminology and parameter names as in previous chapters. In the 2HTM, the model is defined by the following parameters, all of which are defined as probabilities: item parameters D_P, D_R and D_N for performed, read and not presented actions, source parameters d_P and d_R for performed and read actions, item guessing parameter b and source guessing parameters a and g. Both conditions are fitted simultaneously for all participants with parameters defined for both conditions separately.

The total number of parameters exceeds the number of degrees of freedom offered by the data, hence they have to be restricted. First, I used the base structure of the model that best defined the data in Experiment 4. In that experiment, participants responded to performed, observed and novel actions. The best fitting model in that case set the item parameters equal for observed and novel actions, and the source parameters equal for performed and observed actions. Similarly, the model I tested here set item parameters for read and not presented actions equal, and set source parameters for performed and read actions equal. Second, a and g, respectively item guessing parameters for detected and undetected items were set equal. Critically, third, I defined a response bias model that set both item and source memory parameters equal across the observation frequency conditions, allowing only the guessing parameters to vary as an expression of bias.

The response bias model I tested was therefore defined by $D_{P0} = D_{Prep}$, $D_{R0} = D_{Rrep} = D_{N0} = D_{Nrep}$, $d_{P0} = d_{Prep} = d_{R0} = d_{Rrep}$, $a_0 = g_0$, $a_{rep} = g_{rep}$, b_0 and b_{rep}, for a total of seven parameters to account for the 12 degrees of freedom in the data. I fitted the model using the *MPTinR* package (Singmann & Kellen, 2013). The model fit of this model is shown in Table 5.7 in the main body of the thesis. The model provides adequate fit (indicated by a non-significant G^2) for the sum of the individual data with predicted values not significantly different from the observed data. For the aggregate data, predicted and observed data differ significantly. The parameters estimates for this model are given in Table 5.4 in the main body of this thesis.

To test the assumptions of the response bias model, I tested the equality restrictions I set for the model with a series of likelihood ratio tests, i.e., by comparing nested models via comparison of their negative log-likelihoods for the aggregate data. The resulting G^2 value is asymptotically χ^2-distributed with degrees of freedom resulting from the difference in restrictions between the models being compared. I first looked at the item memory parameters. Performed actions led to stronger memories than read or novel actions, $G^2(1) = 30.11, p < .001$, while the strength of read and novel actions did not differ significantly, $G^2(1) = 0.81, p = .37$. Critically, item memory for performed actions, $G^2(1) = 1.87, p = .17$, or read and novel actions, $G^2(1) = 0.18, p = .67$, did not significantly change with repeated observation of those actions.

Second, I looked at the source parameters. There was no significant evidence that the source parameters differed for performed and read actions, $G^2(1) = 0.11, p = .74$. Interestingly, the model that specifies that source memory parameters for performed and read actions increase after repeated observation provided a better fit than the response bias model, $G^2(1) = 9.80, p = .002$. The confidence interval of the source memory parameters in this less restricted model exceeded the range of 0 to 1 (with the upper limit at 1.13). Since the parameters indicate probabilities, interpretation of that range is difficult. I proceeded with the more restricted model that did not allow an increase in source memory parameters but whose parameters remained within-range.

Third, I looked at the guessing parameters. These were the parameters I assumed to change with repeated observation. The item guessing parameter changes with repeated observation of actions, $G^2(1) = 75.98, p < .001$. When participants do not observe the experimenter perform actions after encoding, they are biased to guess an action was not presented, with the parameter clearly below 0.5, $G^2(1) = 67.77, p < .001$. When participants observed actions being performed, on the other hand, they are biased to respond ‘Old’ (‘Performed’ or ‘Read’) to actions, with the parameter clearly exceeding 0.5, $G^2(1) = 18.71, p < .001$. The source guessing parameter equally changes with repeated observation of actions, $G^2(1) = 11.21, p < .001$. When participants do not observe the experimenter perform actions after encoding, they are biased to guess that they had read an action at encoding (versus performed it), with the parameter clearly below 0.5, $G^2(1) = 28.32, p < .001$. On the other hand, if they observed the experimenter perform actions, there is no evidence of biased guessing, with the source guessing parameter not clearly different from 0.5, $G^2(1) = 0.86, p = .35$.

Signal detection models

The basic formulation of the signal detection models follows the ones presented in Chapter 3 and 4. In a signal detection model, items are represented by (typically Gaussian) distributions on one or more memory strength dimensions. Participants make decisions by judging whether items’ strength exceed decision criteria on the dimensions. That means that participants’ decision is based on both items’ strength(s) and the placement of decision criteria. As in previous chapters, I will fit both the two-dimensional and one-dimensional signal detection model.

Two-dimensional signal detection model In the two-dimensional signal detection model (SD-SDT), shown in Figure 3.3 on page 118, each source of memory is represented by a bivariate Gaussian distribution on separable but potentially correlated item memory strength and source memory strength dimensions. Participants judge items to be old (i.e., have been presented at encoding) if the strength of the item exceeds the c_{old} criterion on the item memory strength dimension. In this experiment, this means participants would judge an action to have been read or performed at encoding. Separately, if participants judge an item’s source strength to exceed the $c_{performed}$ criterion, they would respond ‘Performed’, otherwise they would respond ‘Read’. Changes in placement of the criteria reflect changes in participants’ response bias, while changes in placement or shape of the distributions reflects changes in the memory representations.

As with the 2HTM, I tested whether a response bias version of the 2D-SDT could account for the data in Experiment 8. I defined the model as an equal-variance model where the distribution of novel items not presented at encoding had a mean strength of 0 on both item memory and source memory strength dimension. As in previous chapters, I defined the item memory and source memory dimension to be uncorrelated to limit the number of parameters. The mean of the distributions for performed and read actions were defined relative to the baseline of novel actions, but set equal for actions not observed and actions observed repeatedly, for both memory dimensions. The decision criteria c_{old} and $c_{performed}$ were set to vary between observation conditions. This model was therefore defined by a total of 8 parameters: D_P for the item memory strength of performed actions after no observation and repeated observation; D_R for the item memory strength of read actions, after no and repeated observation; d_P for the item memory strength of performed actions after no observation and repeated observation; d_R for the item memory strength of read actions, after no and repeated observation; $c_{old,0}$ for the decision criterion on the item memory strength dimension after no observation and $c_{old,rep}$ for the criterion after repeated observations; $c_{performed,0}$ for the decision criterion on the source memory strength dimension after no observation and $c_{performed,rep}$ for the criterion after repeated observations.

I fitted the model using the *MPTinR* package (Singmann & Kellen, 2013). The modelfit for the response bias model is shown in Table 5.7 in the main body of the thesis. The parameter estimates are given in Table 5.5 in the main body of the thesis. The model does not provide a great fit to the data for either the sum of individual or aggregate data, with the observed data differing significantly from the model restriction indicated by significant G^2 values. As with the previous models, I tested the equality restrictions of this model with likelihood ratio tests. The item memory strength of performed actions, $G^2(1) = 182.65, p < .001$, and read actions, $G^2(1) = 72.79, p < .001$, exceeded that of novel actions. There was no significant evidence that memory for performed actions was stronger than memory for read actions, $G^2(1) = 1.55, p = .21$. I set the memory strength for performed and read items to be equal regardless of items having been observed or not subsequent to encoding. Indeed, there was no evidence that observation changed the item memory strength of performed, $G^2(1) < .001, p > .99$, or read actions, $G^2(1) = .005, p = .94$ in addition to shifts in response bias.

While the source memory strength of performed actions clearly exceeded that of novel actions, $G^2(1) = 103.27, p < .001$, there was no evidence that the source memory strength of read actions did, $G^2(1) < .001, p > .99$. Source memory strength was higher for performed than for read actions, $G^2(1) = 83.78, p < .001$. I set the memory strength for performed and read items to be equal regardless of items having been observed or not subsequent to encoding. Indeed, there was no evidence that observation changed the source memory strength of performed, $G^2(1) < .001, p > .99$, or read actions, $G^2(1) < .001, p = .99$ in addition to shifting response bias. On the other hand, repeated observation lead to more liberal responding by shifting c_{old}, $G^2(1) = 92.29, p < .001$, and $c_{performed}$, $G^2(1) = 11.61, p < .001$.

One-dimensional signal detection model In the one-dimensional model shown in Figure 3.4 on page 120, each source of memory is represented by a Gaussian distribution on a single overall memory strength dimension. Participants judge items to be old (i.e., have been presented at encoding) if the strength of the item exceeds the c_{old} criterion. In this experiment, this means participants would judge an action to have been read or performed at encoding. If an item's strength additionally exceeds the $c_{performed}$ criterion, participants respond 'Performed' otherwise they respond 'Read'. Changes in placement of the criteria reflects changes in participants' response bias, while changes in placement or shape of the distributions reflects changes in the memory representations.

As with the previous models, I tested whether a response bias version of the one-dimensional signal detection model could account for the data in Experiment 8. I defined the model as an equal-variance model where the distribution of novel items not presented at encoding had a mean strength of 0. The mean of the distributions for performed and read actions were defined relative to the baseline of novel actions, but set equal for actions not observed and actions observed repeatedly. The decision criteria c_{old} and $c_{performed}$ were set to vary between observation conditions. This model was therefore defined by a total of 6 parameters: D_P for the memory strength of performed actions after no observation and repeated observation; D_R for the memory strength of read actions, after no and repeated observation; $c_{old,0}$ for the Old/New decision criterion after no observation and $c_{old,rep}$ for the Old/New decision criterion after repeated observation; $c_{performed,0}$ for the Performed/Read decision criterion after no observation and $c_{performed,rep}$ for the Performed/Read decision criterion after repeated observation.

I fitted the model using the *MPTinR* package (Singmann & Kellen, 2013). The modelfit for the response bias model is shown in Table 5.7 in the main body of the thesis. As with the two-dimensional model, the significant G^2 values indicate that that the predicted data

Table D.1. Mean confidence ratings for ‘performed’ responses collapsed across Perspective in Experiment 8

	Source		
Observation frequency	Performed	Read	Novel
Not observed	5.39 (0.73)	5.13 (1.03)	4.93 (1.15)
Observed once	5.44 (0.50)	4.52 (1.16)	4.09 (1.71)
Observed five times	5.44 (0.86)	4.61 (1.27)	4.70 (1.25)

Note. Confidence ratings were made on a scale of 1-Not certain at all to 6-Very certain.

differs significantly from the observed data, suggesting no adequate model fit overall. The parameters estimates for this response bias model are given in Table 5.6 in the main body of the thesis.

I tested whether the equality restrictions hold with this response bias model with a series of likelihood ratio tests. The memory strength of performed actions, $G^2(1) = 279.12, p < .001$, and read actions, $G^2(1) = 48.03, p < .001$, exceeded that of novel actions. Memory for performed actions was stronger than memory for read actions, $G^2(1) = 86.22, p < .001$. I set the memory strength for performed and read items to be equal regardless of items having been observed or not subsequent to encoding. Indeed, there was no evidence that observation changed the memory strength of performed, $G^2(1) = .03, p = .85$, or read actions, $G^2(1) = 2.81, p = .09$ in addition to shifts in response bias. On the other hand, repeated observation lead to more liberal responding by shifting c_{old}, $G^2(1) = 84.87, p < .001$, and $c_{performed}$, $G^2(1) = 46.00, p < .001$.

D.3 Analysis of mean confidence ratings

Participants gave confidence ratings for individual responses on a scale of 1 - not certain at all to 6 - very certain. The mean confidence ratings for ‘performed’ responses are shown in Table D.1. I looked more closely at the mean confidence ratings participants gave when they made a false ‘performed’ response, that is when they responded ‘Performed’ to an action they had only read in the initial encoding phase or had not seen at all. If motor activation in the observer’s motor system during observation leads to false responses, the additional false-source-confirming information should make participants more certain in their false response after observation. If participants falsely respond ‘Performed’ because they lack alternative response options, they should be less confident in their response.

I analysed the data with the *afex* package (Singmann et al., 2015) since not all participants responded ‘Performed’ to all items. I collapsed across read and novel actions and looked only at the effect of Observation frequency. The effect was marginally significant, $F(2,70.96) = 2.50, p = .09$.

While there is a trend for a decrease in confidence in false ‘performed’ responses after observation, the decrease is only statistically significant for actions observed once, $F(1,35.03) = 4.57, p = .04$, and not for actions observed repeatedly, $F(1,43.67) = 1.44, p = .24$. A post-hoc argument could be that the conflict of source evidence (and hence lack of confidence in the source response) is strongest after a single observation, with repeated observations tipping the scale in favour of ‘performed’ responses and resolving the conflict more clearly towards the false response.

Appendix E

Effect sizes of generation and enactment effects in extended recall

Table E.1 shows the proportions of encoded items participants recalled at retrieval in the extended recall task (i.e. how many items participants had available to them at test prior to making source judgements divided by the number of items of that source they encoded). Hollins et al. (2016) asked participants to take turns generating category exemplars. At test, participants were asked to recall own or partner-generated items. In Experiment 3, 5, 6 and 7 in this thesis I asked participants to take turns performing and observing motor actions (Experiment 3) or actions enacted with objects (Experiments 5, 6 and 7). Participants generated all actions they performed themselves. At test, participants were asked to recall self-performed or observed actions.

I calculated Cohen's d_{av} for the difference in proportions of self-generated and heard as well as performed and observed items participants recalled at test, as an approximation of the difference in memory strength between self-generated and heard as well as performed and observed actions.

Table E.1. Proportion of items recalled in extended recall experiments in Hollins et al. (2016)

	Recall own		Recall partner		Mean	
	Own	Partner	Performed	Observed	Performed	Observed
Hollins et al. (2016)[1]						
Prop	0.72 (0.11)	0.45 (0.11)	0.62 (0.16)	0.50 (0.13)	0.67 (0.14)	0.48 (0.12)
d_{av}	2.40		0.79		1.48	
Experiment 3[2]						
Prop	0.49 (0.12)	0.26 (0.11)	0.38 (0.16)	0.36 (0.14)	0.44 (0.14)	0.31 (0.13)
d_{av}	2.01		0.11		0.92	
Experiment 5[3]						
Prop	0.39 (0.11)	0.25 (0.08)	0.33 (0.19)	0.32 (0.13)	0.36 (0.16)	0.29 (0.11)
d_{av}	1.46		0.04		0.55	
Experiment 6[4]						
Prop	0.42 (0.10)	0.25 (0.09)	0.35 (0.10)	0.32 (0.11)	0.38 (0.10)	0.29 (0.10)
d_{av}	1.72		0.27		0.96	
Experiment 7[5]						
Prop	0.42 (0.12)	0.33 (0.15)	0.46 (0.11)	0.32 (0.11)	0.44 (0.12)	0.33 (0.13)
d_{av}	0.35		1.27		0.88	

[1] Hollins et al. (2016), Experiment 3, 1 week delay, verbal items (generated vs heard), Total number of items = 64

[2] Control condition, 1 day delay, motor actions (generated+enacted vs observed), Total number of items = 20

[3] Participant-generated condition, 1 day delay, actions with objects (generated+enacted vs observed), Total number of items = 30

[4] Participant-generated condition, 1 day delay, actions with objects (generated+enacted vs observed), Total number of items = 30

[5] Participant-generated conditions, 1 day delay, actions with objects (self-generated+enacted vs partner-generated+observed), Total number of items = 30

References

Bach, P., Allami, B. K., Tucker, M., & Ellis, R. (2014). Planning-related motor processes underlie mental practice and imitation learning. *Journal of Experimental Psychology: General*, *143*, 1277–1294. doi: 10.1037/a0035604

Bach, P., Bayliss, A. P., & Tipper, S. P. (2011). The predictive mirror: interactions of mirror and affordance processes during action observation. *Psychonomic Bulletin & Review*, *18*, 171–176. doi: 10.3758/s13423-010-0029-x

Bach, P., Fenton-Adams, W., & Tipper, S. P. (2014). Can't touch this: the first-person perspective provides privileged access to predictions of sensory action outcomes. *Journal of Experimental Psychology. Human Perception and Performance*, *40*, 457–464. doi: 10.1037/a0035348

Bach, P., Peatfield, N. A., & Tipper, S. P. (2007). Focusing on body sites: the role of spatial attention in action perception. *Experimental Brain Research*, *178*, 509–517. doi: 10.1007/s00221-006-0756-4

Bach, P., & Tipper, S. P. (2007). Implicit action encoding influences personal-trait judgments. *Cognition*, *102*, 151–178. doi: 10.1016/j.cognition.2005.11.003

Bäckman, L., Nilsson, L.-G., & Chalom, D. (1986). New evidence on the nature of the encoding of action events. *Memory & Cognition*, *14*, 339–346. doi: 10.3758/bf03202512

Bäckman, L., Nilsson, L.-G., & Kormi-Nouri, R. (1993). Attentional demands and recall of verbal and color information in action events. *Scandinavian Journal of Psychology*, *34*, 246–254. doi: 10.1111/j.1467-9450.1993.tb01119.x

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science*, *11*(4), 267–273. (00116)

Batchelder, W. H., & Alexander, G. E. (2013). Discrete-state models: comment on Pazzaglia, Dube, and Rotello (2013). *Psychological Bulletin*, *139*(6), 1204–1212. doi: 10.1037/a0033894

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*, 548–564. doi: 10.1037/0033-295X.97.4.548

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 197–215. doi: 10.1037/0278-7393.22.1.197

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C.-L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition*, *28*, 480–500. doi: 10.3758/BF03198562

Benjamin, A. S., & Craik, F. I. M. (2001). Parallel effects of aging and time pressure on memory for source: Evidence from the spacing effect. *Memory & Cognition*, *29*, 691–697. doi: 10.3758/BF03200471

Berry, C. J., Shanks, D. R., & Henson, R. N. A. (2008). A unitary signal-detection model of implicit and explicit memory. *Trends in Cognitive Sciences*, *12*, 367–373. doi: 10.1016/j.tics.2008.06.005

Berry, C. J., Shanks, D. R., Li, S., Rains, L. S., & Henson, R. N. (2010). Can “pure” implicit memory be isolated? A test of a single-system model of recognition and repetition priming. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *64*(4), 241.

Berry, C. J., Shanks, D. R., Speekenbrink, M., & Henson, R. N. (2012). Models of recognition, repetition priming, and fluency: exploring a new framework. *Psychological review*, *119*, 40–79. doi: 10.1037/a0025464

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*, 201–210. doi: 10.3758/BF03193441

Bink, M. L., Marsh, R. L., & Hicks, J. L. (1999). An alternative conceptualization to memory" strength" in reality monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 804.

Blakemore, S.-J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*(8), 561–567.

Bortoletto, M., Mattingley, J. B., & Cunnington, R. (2013). Effects of Context on Visuomotor Interference Depends on the Perspective of Observed Actions. *PLoS ONE*, *8*. doi: 10.1371/journal.pone.0053248

Bousfield, W. A., & Rosner, S. R. (1970). Free vs unhibited recall. *Psychonomic Science*, *20*(2), 75–76.

Brandimonte, M. A., & Passolunghi, M. C. (1994). The effect of cue-familiarity, cue-distinctiveness, and retention interval on prospective remembering. *The Quarterly Journal of Experimental Psychology*, *47*, 565–587. doi: 10.1080/14640749408401128

Brandt, V. C., Bergström, Z. M., Buda, M., Henson, R. N. A., & Simons, J. S. (2013). Did I turn off the gas? Reality monitoring of everyday actions. *Cognitive, Affective, & Behavioral Neuroscience*, *14*, 209–219. doi: 10.3758/s13415-013-0189-z

Brass, M., Bekkering, H., Wohlschläger, A., & Prinz, W. (2000). Compatibility between Observed and Executed Finger Movements: Comparing Symbolic, Spatial, and Imitative Cues. *Brain and Cognition*, *44*, 124–143. doi: 10.1006/brcg.2000.1225

Bröder, A., & Schütz, J. (2009). Recognition ROCs are curvilinear—or are they? On premature arguments against the two-high-threshold model of recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 587–606. doi: 10.1037/a0015279

Brodeur, M. B., Pelletier, M., & Lepage, M. (2009). Memory for everyday actions in schizophrenia. *Schizophrenia research*, *114*, 71–78. doi: 10.1016/j.schres.2009.06.023

Brooks, B. M., & Gardiner, J. M. (1994). Age differences in memory for prospective compared with retrospective subject-performed tasks. *Memory & Cognition*, *22*, 27–33. doi: 10.3758/BF03202758

Brown, A. S., & Murphy, D. R. (1989). Cryptomnesia: Delineating inadvertent plagiarism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 432–442. doi: 10.1037/0278-7393.15.3.432

Chan, J. C., Wilford, M. M., & Hughes, K. L. (2012). Retrieval can increase or decrease suggestibility depending on how memory is tested: The importance of source complexity. *Journal of Memory and Language*, *67*, 78–85. doi: 10.1016/j.jml.2012.02.006

Chong, T. T. J., Cunnington, R., Williams, M. A., Kanwisher, N., & Mattingley, J. B. (2008). fMRI Adaptation Reveals Mirror Neurons in Human Inferior Parietal Cortex. *Current Biology*, *18*, 1576–1580. doi: 10.1016/j.cub.2008.08.068

Cohen, A. L., Sanborn, A. N., & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review*, *15*, 692–712. doi: 10.3758/PBR.15.4.692

Cohen, G., & Faulkner, D. (1989). Age differences in source forgetting: Effects on reality monitoring and on eyewitness testimony. *Psychology and Aging*, *4*, 10–17. (00299) doi: 10.1037/0882-7974.4.1.10

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology*, *22*, 267–281. doi: 10.1111/j.1467-9450.1981.tb00402.x

Cohen, R. L. (1983). The effect of encoding variables on the free recall of words and action events. *Memory & Cognition*, *11*, 575–582. doi: 10.3758/BF03198282

Cohen, R. L. (1988). Metamemory for words and enacted instructions: Predicting which items will be recalled. *Memory & Cognition*, *16*, 452–460. doi: 10.3758/BF03214226

Conway, M. A., & Dewhurst, S. A. (1995). Remembering, Familiarity, and Source Monitoring. *The Quarterly Journal of Experimental Psychology Section A*, *48*, 125–140. doi: 10.1080/14640749508401380

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of memory and language*, *63*, 465–475. doi: 10.1016/j.jml.2010.07.002

Craighero, L., Bello, A., Fadiga, L., & Rizzolatti, G. (2002). Hand action preparation influences the responses to hand pictures. *Neuropsychologia*, *40*, 492–502. doi: 10.1016/s0028-3932(01)00134-8

Craighero, L., Fadiga, L., Rizzolatti, G., & Umiltà, C. (1999). Action for perception: a motor-visual attentional effect. *Journal of experimental psychology: Human perception and performance*, *25*, 1673–1692. doi: 10.1037/0096-1523.25.6.1673

Craik, F. I. M. (2014). Effects of distraction on memory and cognition: a commentary. *Frontiers in Psychology*, *5*. doi: 10.3389/fpsyg.2014.00841

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*, 159–180. doi: 10.1037/0096-3445.125.2.159

Craik, F. I. M., Luo, L., & Sakuta, Y. (2010). Effects of aging and divided attention on memory for items and their contexts. *Psychology and Aging*, *25*, 968–979. doi: 10.1037/a0020276

Csibra, G. (2007). Action mirroring and action understanding: an alternative account. In P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorymotor Foundations of Higher Cognition. Attention and Performance XXII* (pp. 435–459). Oxford: Oxford University Press.

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* London: Routledge.

DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology*, *47*, 292–303. doi: 10.1016/s0022-2496(03)00005-1

Defeldre, A.-C. (2005). Inadvertent plagiarism in everyday life. *Applied Cognitive Psychology*, *19*, 1033–1040. doi: 10.1002/acp.1129

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Palgrave Macmillan.

Dodson, C. S., Holland, P. W., & Shimamura, A. P. (1998). On the recollection of specific- and partial-source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1121–1136. doi: 10.1037/0278-7393.24.5.1121

Dodson, C. S., & Johnson, M. K. (1993). Rate of false source attributions depends on how questions are asked. *The American Journal of Psychology*, *106*, 541–557. doi: 10.2307/1422968

Dodson, C. S., & Johnson, M. K. (1996). Some problems with the process-dissociation approach to memory. *Journal of Experimental Psychology: General*, *125*, 181–194. doi: 10.1037/0096-3445.125.2.181

Dodson, C. S., & Schacter, D. L. (2001). “If I had said it I would have remembered it: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*, 155–161. doi: 10.3758/bf03196152

Drivdahl, S. B., & Zaragoza, M. S. (2001). The role of perceptual elaboration and individual differences in the creation of false memories for suggested events. *Applied Cognitive Psychology*, *15*, 265–281. doi: 10.1002/acp.701

Dube, C. (in press). The statistical accuracy and theoretical status of discrete-state MPT models: Reply to Batchelder and Alexander (in press).

Durso, F. T., & Johnson, M. K. (1980). The effects of orienting tasks on recognition, recall, and modality confusion of pictures and words. *Journal of Verbal Learning and Verbal Behavior*, *19*, 416–429. doi: 10.1016/s0022-5371(80)90294-7

Engelkamp, J. (1998). *Memory for actions*. Psychology Press.

Engelkamp, J. (2001). Action memory: a system-oriented approach. In R. L. Cohen, M. J. Guynn, J. Engelkamp, R. Kormi-Nouri, M. A. Foley, & H. D. Zimmer (Eds.), *Memory for Action. A Distinct Form of Episodic Memory?* (pp. 97–111). Oxford: Oxford University Press.

Engelkamp, J., & Dehn, D. M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 671–682. doi: 10.1037//0278-7393.26.3.671

Engelkamp, J., & Krumnacker, H. (1980). Imaginale und motorische Prozesse beim Behalten verbalen Materials. [Image- and motor-processes in the retention of verbal materials.]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *27*, 511–533.

Engelkamp, J., & Zimmer, H. D. (1995). Similarity of movement in recognition of self-performed tasks and of verbal tasks. *British Journal of Psychology*, *86*, 241–252. doi: 10.1111/j.2044-8295.1995.tb02559.x

Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica*, *96*, 43–60. doi: 10.1016/S0001-6918(97)00005-X

Engelkamp, J., Zimmer, H. D., & Biegelmann, U. E. (1993). Bizarreness effects in verbal tasks and subject-performed tasks. *European Journal of Cognitive Psychology*, *5*, 393–415. doi: 10.1080/09541449308520127

Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition*, *22*, 34–39. doi: 10.3758/BF03202759

Eschen, A., Freeman, J., Dietrich, T., Martin, M., Ellis, J., Martin, E., & Kliegel, M. (2007). Motor brain regions are involved in the encoding of delayed intentions: A fMRI study. *International Journal of Psychophysiology*, *64*, 259–268. doi: 10.1016/j.ijpsycho.2006.09.005

Fabbri-Destro, M., & Rizzolatti, G. (2008). Mirror neurons and mirror systems in monkeys and humans. *Physiology (Bethesda, Md.)*, *23*, 171–179. doi: 10.1152/physiol.00004.2008

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*, 175–191. doi: 10.3758/bf03193146

Ferguson, S. A., Hashtroudi, S., & Johnson, M. K. (1992). Age differences in using source-relevant cues. *Psychology and aging*, *7*, 443–452. doi: 10.1037/0882-7974

.7.3.443

Foley, M. A., Bays, R. B., Foy, J., & Woodfield, M. (2015). Source misattributions and false recognition errors: Examining the role of perceptual resemblance and imagery generation processes. *Memory*, *23*, 714–735. doi: 10.1080/09658211.2014.925565

Foley, M. A., Bouffard, V., Raag, T., & DiSanto-Rose, M. (1991). The effects of enactive encoding, type of movement, and imagined perspective on memory of dance. *Psychological Research*, *53*, 251–259. doi: 10.1007/BF00941395

Foley, M. A., & Johnson, M. K. (1985). Confusions between memories for performed and imagined actions: A developmental comparison. *Child development*, 1145–1155. doi: 10.2307/1130229

Foley, M. A., Passalacqua, C., & Ratner, H. H. (1993). Appropriating the actions of another: Implications for children's memory and learning. *Cognitive Development*, *8*, 373–401. doi: 10.1016/S0885-2014(05)80001-2

Freeman, J. E., & Ellis, J. A. (2003). The representation of delayed intentions: A prospective subject-performed task? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 976–992. doi: 10.1037/0278-7393.29.5.976

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the cognitive sciences*, *4*, 23–48. doi: 10.1007/s11097-005-4737-z

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609. doi: 10.1093/brain/119.2.593

Gallivan, J. P., McLean, D. A., Valyear, K. F., Pettypiece, C. E., & Culham, J. C. (2011). Decoding Action Intentions from Preparatory Brain Activity in Human Parieto-Frontal Networks. *The Journal of Neuroscience*, *31*, 9599–9610. doi: 10.1523/JNEUROSCI.0080-11.2011

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review*, *3*, 208–214. doi: 10.3758/bf03212420

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2016). *mvtnorm: Multivariate Normal and t Distributions.* Retrieved from `http://CRAN.R-project.org/package=mvtnorm` (R package version 1.0-5)

Gingerich, A. C., & Sullivan, M. C. (2013). Claiming hidden memories as one's own: A review of inadvertent plagiarism. *Journal of Cognitive Psychology*, *25*, 903–916.

(00003) doi: 10.1080/20445911.2013.841674

Goff, L. M., & Roediger, H. L. (1998). Imagination inflation for action events: repeated imaginings lead to illusory recollections. *Memory & Cognition*, *26*, 20–33. doi: 10.3758/bf03211367

Goldsmith, M. (2016). Metacognitive Quality-Control Processes in Memory Retrieval and Reporting. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory.* Oxford: Oxford University Press.

Golly-Haring, C., & Engelkamp, J. (2003). Categorical-relational and order-relational information in memory for subject-performed and experimenter-performed actions. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *29*(5), 965–975. doi: 10.1037/0278-7393.29.5.965

Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Human brain mapping*, *12*, 1–19.

Gutsell, J. N., & Inzlicht, M. (2010). Empathy constrained: Prejudice predicts reduced mental simulation of actions during observation of outgroups. *Journal of Experimental Social Psychology*, *46*, 841–845. doi: 10.1016/j.jesp.2010.03.011

Guzel, M. A., & Higham, P. A. (2013). Dissociating early- and late-selection processes in recall: The mixed blessing of categorized study lists. *Memory & Cognition*, *41*, 683–697. doi: 10.3758/s13421-012-0292-3

Halamish, V., Goldsmith, M., & Jacoby, L. L. (2012). Source-constrained recall: Front-end and back-end control of retrieval quality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1–15. doi: 10.1037/a0025053

Hamilton, A., Wolpert, D., & Frith, U. (2004). Your own action influences how you perceive another person's action. *Current Biology*, *14*, 493–498. doi: 10.1016/j.cub.2004.03.007

Hanakawa, T., Dimyan, M. A., & Hallett, M. (2008). Motor Planning, Imagery, and Execution in the Distributed Motor Network: A Time-Course Study with Functional MRI. *Cerebral Cortex*, *18*, 2775–2788. doi: 10.1093/cercor/bhn036

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology and aging*, *5*, 119–126. (00294) doi: 10.1037/0882-7974.5.1.119

Hautus, M. J., Macmillan, N. A., & Rotello, C. B. (2008). Toward a complete decision model of item and source recognition. *Psychonomic Bulletin & Review*, *15*, 889–905. doi: 10.3758/PBR.15.5.889

Helstrup, T. (1987). One, two, or three memories? A problem-solving approach to memory for performed acts. *Acta Psychologica*, *66*, 37–68. doi: 10.1016/0001-6918(87)90017-5

Helstrup, T. (1999). Visuo-spatial encoding of movement patterns. *European Journal of Cognitive Psychology*, *11*, 357–371. doi: 10.1080/713752325

Helstrup, T. (2001). Concurrent and retroactive interference effects in memory of movement patterns. *The Quarterly Journal of Experimental Psychology Section A*, *54*, 547–560. doi: 10.1080/713755979

Helstrup, T. (2005). In search of a motor element in memory for enacted events. *European Journal of Cognitive Psychology*, *17*, 389–403. doi: 10.1080/09541440440000087

Henkel, L. A. (2011). Photograph-induced memory errors: When photographs make people claim they have done things they have not. *Applied Cognitive Psychology*, *25*, 78–86. doi: 10.1002/acp.1644

Henkel, L. A., & Carbuto, M. (2008). Remembering what we did: How source misattributions arise from verbalization, mental imagery, and pictures. In M. R. Kelley (Ed.), *Applied Memory* (pp. 213 – 234). New York: Nova Science Publishers.

Hesse, C., & Franz, V. H. (2009). Memory mechanisms in grasping. *Neuropsychologia*, *47*, 1532–1545. doi: 10.1016/j.neuropsychologia.2008.08.012

Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). Task interference in time-based, event-based, and dual intention prospective memory conditions. *Journal of Memory and Language*, *53*, 430–444. doi: 10.1016/j.jml.2005.04.001

Hoffman, H. G. (1997). Role of memory strength in reality monitoring decisions: Evidence from source attribution biases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 371–383. doi: 10.1037/0278-7393.23.2.371

Hollins, T. J., Lange, N., Berry, C. J., & Dennis, I. (2016). Giving and stealing ideas in memory: Source errors in recall are influenced by both early-selection and late-correction retrieval processes. *Journal of Memory and Language*, *88*, 87–103. doi: 10.1016/j.jml.2016.01.004

Hollins, T. J., Lange, N., Dennis, I., & Longmore, C. A. (2015). Social influences on unconscious plagiarism and anti-plagiarism. *Memory*, 1–19. doi: 10.1080/09658211.2015.1059857

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and brain sciences*, *24*, 849–878. doi: 10.1017/s0140525x01000103

Hornstein, S. L., & Mulligan, N. W. (2004). Memory for actions: Enactment and source memory. *Psychonomic Bulletin & Review*, *11*, 367–372. doi: 10.3758/BF03196584

Huang, G. B., Jain, V., & Learned-Miller, E. (2007). Unsupervised joint alignment of complex images. In *Computer Vision (ICCV), 2007 IEEE 11th International Conference on* (pp. 1–8). Rio de Janeiro, Brazil.

Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007, October). *Labeled faces in the wild: A database for studying face recognition in unconstrained environments* (Tech. Rep. No. 07-49). University of Massachusetts, Amherst. Retrieved from `https://www.cs.umass.edu/~elm/papers/lfw.pdf`

Hunt, R. R. (2012). Distinctive Processing. The co-action of similarity and difference in memory. *Psychology of Learning and Motivation - Advances in Research and Theory*, *56*, 1–46. doi: 10.1016/B978-0-12-394393-4.00001-7

Iacoboni, M. (2005). Understanding others: Imitation, language, empathy. *Perspectives on imitation: From cognitive neuroscience to social science*, *1*, 77–99.

Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M.-C., ... Rizzolatti, G. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proceedings of the national academy of sciences*, *98*, 13995–13999.

Ikier, S., Tekcan, A. İ., Gülgöz, S., & Küntay, A. C. (2003). Whose life is it anyway? Adoption of each other's autobiographical memories by twins. *Applied Cognitive Psychology*, *17*, 237–247. doi: 10.1002/acp.869

Ikkai, A., & Curtis, C. E. (2011). Common neural mechanisms supporting spatial working memory, attention and motor intention. *Neuropsychologia*, *49*, 1428–1434.

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2006). Neural circuits involved in imitation and perspective-taking. *NeuroImage*, *31*, 429–439. doi: 10.1016/j.neuroimage.2005.11.026

Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of personality and social psychology*, *56*, 326–338. doi: 10.1037/0022-3514.56.3.326

Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The Role of Cognitive Control. In

S. Chaiken & Trope (Eds.), *Dual-process theories in social psychology* (pp. 383–400). New York: Guilford.

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of experimental psychology: General*, *118*, 115–125. doi: 10.1037/0096-3445.118.2.115

Jang, Y., Wixted, J. T., & Huber, D. E. (2011). The diagnosticity of individual data for model selection: Comparing signal-detection models of recognition memory. *Psychonomic Bulletin & Review*, *18*, 751–757. doi: 10.3758/s13423-011-0096-7

Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage*, *14*, S103–S109. doi: 10.1006/nimg.2001.0832

Jeannerod, M., & Anquetil, T. (2008). Putting oneself in the perspective of the other: A framework for self–other differentiation. *Social Neuroscience*, *3*, 356–367. doi: 10.1080/17470910701563715

Jeffreys, H. (1961). *The theory of probability*. Oxford University Press. (00000)

Johnson, M. K. (1985). The origin of memories. In P. Kendall (Ed.), *Advances in cognitive-behavioral research and therapy* (Vol. 4, pp. 1–27). New York: Academic Press.

Johnson, M. K., Foley, M. A., & Leach, K. (1988). The consequences for memory of imagining in another person's voice. *Memory & Cognition*, *16*, 337–342. doi: 10.3758/bf03197044

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28. doi: 10.1037/0033-2909.114.1.3

Johnson, M. K., Nolde, S. F., & De Leonardis, D. M. (1996). Emotional focus and source monitoring. *Journal of Memory and Language*, *35*, 135–156. (00171) doi: 10.1006/jmla.1996.0008

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67–85. doi: 10.1037/0033-295X.88.1.67

Kahana, M. J., Dolan, E. D., Sauder, C. L., & Wingfield, A. (2005). Intrusions in Episodic Recall: Age Differences in Editing of Overt Responses. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*, 92–97. doi: 10.1093/geronb/60.2.P92

Kellen, D., & Klauer, K. C. (2015). Signal detection and threshold modeling of confidence-rating ROCs: A critical test with minimal assumptions. *Psychological Review*, *122*, 542–557. doi: 10.1037/a0039251

Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology*, *13*, 522–525. doi: 10.1016/s0960-9822(03)00165-9

Koriat, A., Ben-Zur, H., & Druch, A. (1991). The contextualization of input and output events in memory. *Psychological Research*, *53*, 260–270. doi: 10.1007/BF00941396

Koriat, A., Ben-Zur, H., & Nussbaum, A. (1990). Encoding information for future action: Memory for to-be-performed tasks versus memory for to-be-recalled tasks. *Memory & Cognition*, *18*, 568–578. doi: 10.3758/bf03197099

Kormi-Nouri, R. (2000). The role of movement and object in action memory: a comparative study between blind, blindfolded and sighted subjects. *Scandinavian journal of psychology*, *41*, 71–76. doi: 10.1111/1467-9450.00173

Kormi-Nouri, R., & Nilsson, L.-G. (1998). The role of integration in recognition failure and action memory. *Memory & cognition*, *26*, 681–691. doi: 10.3758/bf03211389

Kormi-Nouri, R., Nyberg, L., & Nilsson, L.-G. (1994). The effect of retrieval enactment on recall of subject-performed tasks and verbal tasks. *Memory & Cognition*, *22*, 723–728. doi: 10.3758/BF03209257

Koski, L. M., Iacoboni, M., Dubeau, M.-C., Woods, R. P., & Mazziotta, J. C. (2001). Human area po? Visuomotor transformations in parieto-occipital cortex during imitation. *NeuroImage*, *13*, 1204. doi: 10.1016/S1053-8119(01)92521-2

Krakauer, J. W., & Shadmehr, R. (2006). Consolidation of motor memory. *Trends in Neurosciences*, *29*, 58–64. doi: 10.1016/j.tins.2005.10.003

Küntay, A. C., Gülgöz, S., & Tekcan, A. İ. (2004). Disputed memories of twins: how ordinary are they? *Applied Cognitive Psychology*, *18*, 405–413. doi: 10.1002/acp.976

Lampinen, J. M., Odegard, T. N., & Bullington, J. L. (2003). Qualities of memories for performed and imagined actions. *Applied Cognitive Psychology*, *17*, 881–893. doi: 10.1002/acp.916

Landau, J. D., & Marsh, R. L. (1997). Monitoring source in an unconscious plagiarism paradigm. *Psychonomic Bulletin & Review*, *4*, 265–270. doi: 10.3758/BF03209404

Lane, S. M. (2006). Dividing attention during a witnessed event increases eyewitness suggestibility. *Applied Cognitive Psychology*, *20*, 199–212. doi: 10.1002/acp.1177

Lewandowsky, S., & Farrell, S. (2011). *Computational Modeling in Cognition: Principles and Practice*. SAGE Publications.

Leynes, P. A., & Kakadia, B. (2013). Variations in retrieval monitoring during action memory judgments: Evidence from event-related potentials (ERPs). *International Journal of Psychophysiology*, *87*, 189–199. doi: 10.1016/j.ijpsycho.2013.01.004

Lichty, W., Bressie, S., & Krell, R. (1988). When a fork is not a fork: Recall of performed activities as a function of age, generation and bizarreness. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues.* (pp. 506–511). Chichester: Wiley.

Lindner, I., & Davidson, P. S. (2014). False action memories in older adults: Relationship with executive functions? *Aging, Neuropsychology, and Cognition*, *21*, 560–576. doi: 10.1080/13825585.2013.839026

Lindner, I., & Echterhoff, G. (2015). Imagination inflation in the mirror: Can imagining others' actions induce false memories of self-performance? *Acta Psychologica*, *158*, 51–60. doi: 10.1016/j.actpsy.2015.03.008

Lindner, I., Echterhoff, G., Davidson, P. S. R., & Brand, M. (2010). Observation Inflation Your Actions Become Mine. *Psychological Science*, *21*, 1291–1299. doi: 10.1177/0956797610379860

Lindner, I., & Henkel, L. A. (2015). Confusing what you heard with what you did: False action-memories from auditory cues. *Psychonomic Bulletin & Review*(22), 1791–1797. doi: 10.3758/s13423-015-0837-0

Lindner, I., Schain, C., & Echterhoff, G. (2016). Other-self confusions in action memory: The role of motor processes. *Cognition*, *149*, 67–76. (00000) doi: 10.1016/j.cognition.2016.01.003

Lindner, I., Schain, C., Kopietz, R., & Echterhoff, G. (2012). When Do We Confuse Self and Other in Action Memory? Reduced False Memories of Self-Performance after Observing Actions by an Out-Group vs. In-Group Actor. *Frontiers in Psychology*, *3*. doi: 10.3389/fpsyg.2012.00467

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology*, *52*, 297–318. doi: 10.1016/0022-0965(91)90065-Z

Lockhart, R. S., & Craik, F. I. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *44*, 87–112. doi: 10.1037/h0084237

Lyle, K., & Johnson, M. (2006). Importing perceived features into false memories. *Memory*, *14*, 197–213. doi: 10.1080/09658210544000060

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology Press.

Macrae, C. N., Bodenhausen, G. V., & Calvini, G. (1999). Contexts of cryptomnesia: May the source be with you. *Social Cognition*, *17*(3), 273–297. doi: 10.1521/soco.1999.17.3.273

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*, 49–70. doi: 10.3758/BF03211715

Manzi, A., & Nigro, G. (2008). Long-term memory for performed and observed actions: Retrieval awareness and source monitoring. *Memory*, *16*, 595–603. doi: 10.1080/09658210802070749

Marsh, R. L., & Bower, G. H. (1993). Eliciting cryptomnesia: unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *19*, 673–688. doi: 10.1037/0278-7393.19.3.673

Marsh, R. L., & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1137–1151. doi: 10.1037/0278-7393.24.5.1137

Marsh, R. L., & Landau, J. D. (1995). Item availability in cryptomnesia: Assessing its role in two paradigms of unconscious plagiarism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1568–1582. doi: 10.1037/0278-7393.21.6.1568

Marsh, R. L., Landau, J. D., & Hicks, J. L. (1997). Contributions of inadequate source monitoring to unconscious plagiarism during idea generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 886–897. doi: 10.1037/0278-7393.23.4.886

Masumoto, K., Yamaguchi, M., Sutani, K., Tsuneto, S., Fujita, A., & Tonoike, M. (2006). Reactivation of physical motor information in the memory of action events. *Brain Research*, *1101*, 102–109. doi: 10.1016/j.brainres.2006.05.033

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, *114*, 1–16. doi: 10.1037/0096-3445.114.1.1

McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: a common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*, *15*, 237–255. doi: 10.3758/pbr.15.2.237

McDaniel, M. A., Lyle, K. B., Butler, K. M., & Dornburg, C. C. (2008). Age-related deficits in reality monitoring of action memories. *Psychology and Aging*, *23*, 646–656. doi: 10.1037/a0013083

McDermott, K. B., & Roediger, H. L. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language*, *39*, 508–520. doi: 10.1006/jmla.1998.2582

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review*, *14*, 858–865. doi: 10.3758/BF03194112

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. In *The Oxford handbook of memory* (pp. 179–195).

Morey, R. D., Pratte, M. S., & Rouder, J. N. (2008). Problematic effects of aggregation in z ROC analysis and a hierarchical modeling solution. *Journal of Mathematical Psychology*, *52*, 376–388. (00036) doi: 10.1016/j.jmp.2008.02.001

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes Factors for Common Designs.* Retrieved from `http://CRAN.R-project.org/package=BayesFactor`

Mulligan, N. W. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 838–855. doi: 10.1037/0278-7393.30.4.838

Mulligan, N. W., Besken, M., & Peterson, D. (2010). Remember-Know and source memory instructions can qualitatively change old-new recognition accuracy: the modality-match effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 558–566. doi: 10.1037/e520562012-257

Mulligan, N. W., & Hornstein, S. L. (2003). Memory for actions: self-performed tasks and the reenactment effect. *Memory & Cognition*, *31*, 412–421. doi: 10.3758/

bf03194399

Murnane, K., & Bayen, U. J. (1996, July). An evaluation of empirical measures of source identification. *Memory & Cognition*, *24*(4), 417–428. doi: 10.3758/BF03200931

Nilsson, L.-G. (2000). Remembering actions and words. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 137–148). Oxford: Oxford University Press. (00000)

Nilsson, L.-G., & Bäckman, L. (1989). Implicit memory and the enactment of verbal instructions. In S. Lewandowsky, J. C. Dunn, & K. Kirsner (Eds.), *Implicit Memory: Theoretical Issues* (pp. 173–183). Hillsdale, NJ: Erlbaum.

Nilsson, L.-G., & Cohen, R. L. (1988). Enrichment and generation in the recall of enacted and non-enacted instructions. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues, Vol. 1: Memory in everyday life* (pp. 427–432). Oxford, England: John Wiley & Sons.

Nilsson, L.-G., Nyberg, L., Nouri, R. K., & Rönnlund, M. (1995). Dissociative effects of elaboration on memory of enacted and non-enacted events: A case of a negative effect. *Scandinavian Journal of Psychology*, *36*, 225–231. doi: 10.1111/j.1467-9450.1995.tb00981.x

Norris, M. P., & West, R. L. (1993). Activity memory and aging: The role of motor retrieval and strategic processing. *Psychology and Aging*, *8*, 81–86. doi: 10.1037/0882-7974.8.1.81

Nosofsky, R. M., & Kruschke, J. K. (2001). Single-system models and interference in category learning: Commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review*, *9*, 169–174. doi: 10.3758/BF03196274

Nyberg, L., Petersson, K. M., Nilsson, L.-G., Sandblom, J., Åberg, C., & Ingvar, M. (2001). Reactivation of Motor Brain Areas during Explicit Memory for Actions. *NeuroImage*, *14*, 521–528. doi: 10.1006/nimg.2001.0801

Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2013). Crossmodal and action-specific: neuroimaging the human mirror neuron system. *Trends in cognitive sciences*, *17*, 311–318. doi: 10.1016/j.tics.2013.04.012

Oosterhof, N. N., Wiggett, A. J., Diedrichsen, J., Tipper, S. P., & Downing, P. E. (2010). Surface-based information mapping reveals crossmodal vision–action representations in human parietal and occipitotemporal cortex. *Journal of Neurophysiology*, *104*, 1077–1089. doi: 10.1152/jn.00326.2010

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.

Pazzaglia, A. M., Dube, C., & Rotello, C. M. (2013). A critical comparison of discrete-state and continuous models of recognition memory: Implications for recognition and beyond. *Psychological Bulletin*, *139*, 1173–1203. doi: 10.1037/a0033044

Perfect, T. J., Field, I., & Jones, R. (2009). Source credibility and idea improvement have independent effects on unconscious plagiarism errors in recall and generate-new tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 267–274. doi: 10.1037/a0013936

Perfect, T. J., & Stark, L. J. (2008a). Tales from the Crypt... omnesia. In J. Dunlosky & Bjork (Eds.), *A handbook of metamemory and memory* (pp. 285–314). New York: NY: LEA.

Perfect, T. J., & Stark, L.-J. (2008b). Why do I always have the best ideas? The role of idea quality in unconscious plagiarism. *Memory (Hove, England)*, *16*, 386–394. doi: 10.1080/09658210801946501

Pfeifer, J. H., Iacoboni, M., Mazziotta, J. C., & Dapretto, M. (2008). Mirroring others' emotions relates to empathy and interpersonal competence in children. *Neuroimage*, *39*, 2076–2085. doi: 10.1016/j.neuroimage.2007.10.032

Press, C., Bird, G., Walsh, E., & Heyes, C. (2008). Automatic imitation of intransitive actions. *Brain and Cognition*, *67*, 44–50. doi: 10.1016/j.bandc.2007.11.001

Press, C., Cook, J., Blakemore, S.-J., & Kilner, J. (2011). Dynamic modulation of human motor activity when observing actions. *The Journal of Neuroscience*, *31*, 2792–2800. doi: 10.1523/jneurosci.1595-10.2011

Raaijmakers, J. G., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. *Psychology of Learning and Motivation*, *14*, 207–262. doi: 10.1016/s0079-7421(08)60162-0

Ratner, H. H., Foley, M. A., & Gimpert, N. (2000). Person perspectives on children's memory and learning: What do source-monitoring failures reveal. In K. P. Roberts & M. Blades (Eds.), *Children's Source Monitoring* (pp. 85–114). Psychology Press.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*, 318–339. doi: 10.1037/0033-295x.95.3.318

Riefer, D. M., Hu, X., & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 680–693. doi: 10.1037/0278-7393.20.3.680

Rizzolatti, G., & Craighero, L. (2004). The Mirror-Neuron System. *Annual Review of Neuroscience*, *27*, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., & Fabbri-Destro, M. (2008). The mirror system and its role in social cognition. *Current opinion in neurobiology*, *18*, 179–184. doi: 10.1016/j.conb.2008.08.001

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive brain research*, *3*, 131–141. doi: 10.1016/0926-6410(95)00038-0

Roediger III, H. L., & Zaromb, F. M. (2010). Memory for actions: How different? In *Memory, aging and the brain: A Festschrift in honour of Lars-Göran Nilsson* (pp. 24–52). New York, NY, US: Psychology Press.

Rosa, N. M., Deason, R. G., Budson, A. E., & Gutchess, A. H. (2014). Source Memory for Self and Other in Patients With Mild Cognitive Impairment due to Alzheimer's Disease. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 59–65. doi: 10.1093/geronb/gbu062

Rosa, N. M., & Gutchess, A. H. (2011). Source memory for action in young and older adults: Self vs. close or unknown others. *Psychology and Aging*, *26*, 625–630. doi: 10.1037/a0022827

Saltz, E., & Dixon, D. (1982). Let's pretend: The role of motoric imagery in memory for sentences and words. *Journal of Experimental Child Psychology*, *34*, 77–92. doi: 10.1016/0022-0965(82)90032-7

Saltz, E., & Donnenwerth-Nolan, S. (1981). Does motoric imagery facilitate memory for sentences? A selective interference test. *Journal of Verbal Learning and Verbal Behavior*, *20*, 322–332. doi: 10.1016/S0022-5371(81)90472-2

Saxe, R., Jamal, N., & Powell, L. (2006). My Body or Yours? The Effect of Visual Perspective on Cortical Body Representations. *Cerebral Cortex*, *16*, 178–182. doi: 10.1093/cercor/bhi095

Schaaf, M. (1988). Motorische Aktivitaet und verbale Lernleistung - Leistungssteigerung durch Simultanitaet? *Zeitschrift ftir experimentelle und angewandte Psychologie*, 298 – 302.

Schain, C., Lindner, I., Beck, F., & Echterhoff, G. (2012). Looking at the actor's face: Identity cues and attentional focus in false memories of action performance from observation. *Journal of Experimental Social Psychology*, *48*, 1201–1204. doi: 10.1016/j.jesp.2012.04.003

Schult, J. C., von Stülpnagel, R., & Steffens, M. C. (2014). Enactment versus Observation: Item-Specific and Relational Processing in Goal-Directed Action Sequences (and Lists of Single Actions). *PloS one*, *9*. doi: 10.1371/journal.pone.0099985

Schütz, J., & Bröder, A. (2011). Signal detection and threshold models of source memory. *Experimental Psychology*, *58*, 293–311. doi: 10.1027/1618-3169/a000097

Seamon, J. G., Philbin, M. M., & Harrison, L. G. (2006). Do you remember proposing marriage to the Pepsi machine? False recollections from a campus walk. *Psychonomic Bulletin & Review*, *13*, 752–756. doi: 10.3758/BF03193992

Senkfor, A. J. (2008). Memory for pantomimed actions versus actions with real objects. *Cortex*, *44*, 820–833. doi: 10.1016/j.cortex.2007.03.003

Senkfor, A. J., Petten, C. V., & Kutas, M. (2002). Episodic Action Memory for Real Objects: An ERP Investigation With Perform, Watch, and Imagine Action Encoding Tasks Versus a Non-Action Encoding Task. *Journal of Cognitive Neuroscience*, *14*, 402–419. doi: 10.1162/089892902317361921

Shanks, D. R., & Berry, C. J. (2012). Are there multiple memory systems? Tests of models of implicit and explicit memory. *The Quarterly Journal of Experimental Psychology*, *65*, 1449–1474. doi: 10.1080/17470218.2012.691887

Sheen, M., Kemp, S., & Rubin, D. (2001). Twins dispute memory ownership: A new false memory phenomenon. *Memory & Cognition*, *29*, 779–788.

Singmann, H., Bolker, B., & Westfall, J. (2015). *afex: Analysis of Factorial Experiments* (Tech. Rep.). Retrieved from `http://CRAN.R-project.org/package=afex` (R package version 0.15-2)

Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models in R. *Behavior Research Methods*, *45*, 560–575. (00000) doi: 10.3758/s13428-012-0259-0

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592–604. doi: 10.1037/0278-7393.4.6.592

Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition*, *33*, 151–170. doi: 10.3758/BF03195305

Slotnick, S. D., Jeye, B. M., & Dodson, C. S. (2016). Recollection is a continuous process: Evidence from plurality memory receiver operating characteristics. *Memory*, *24*, 2–11. doi: 10.1080/09658211.2014.971033

Smyth, M. M., Pearson, N. A., & Pendleton, L. R. (1988). Movement and working memory: Patterns and positions in space. *The Quarterly Journal of Experimental Psychology Section A*, *40*, 497–514. doi: 10.1080/02724988843000041

Stark, L.-J., Perfect, T., & Newstead, S. (2005). When elaboration leads to appropriation: Unconscious plagiarism in a creative task. *Memory*, *13*, 561–573. doi: 10.1080/09658210444000232

Stark, L.-J., & Perfect, T. J. (2006). Elaboration inflation: how your ideas become mine. *Applied Cognitive Psychology*, *20*, 641–648. doi: 10.1002/acp.1216

Stark, L.-J., & Perfect, T. J. (2007). Whose idea was that? Source monitoring for idea ownership following elaboration. *Memory*, *15*, 776–783. doi: 10.1080/09658210701643042

Stefan, K., Cohen, L. G., Duque, J., Mazzocchio, R., Celnik, P., Sawaki, L., … Classen, J. (2005). Formation of a motor memory by action observation. *The Journal of Neuroscience*, *25*, 9339–9346. doi: 10.1523/jneurosci.2282-05.2005

Steffens, M. C. (2007). Memory for goal-directed sequences of actions: Is doing better than seeing? *Psychonomic bulletin & review*, *14*, 1194–1198. doi: 10.3758/bf03193112

Steffens, M. C., Buchner, A., & Wender, K. F. (2003). Quite ordinary retrieval cues may determine free recall of actions. *Journal of Memory and Language*, *48*, 399–415. doi: 10.1016/s0749-596x(02)00517-x

Steffens, M. C., Buchner, A., Wender, K. F., & Decker, C. (2007). Limits on the role of retrieval cues in memory for actions: Enactment effects in the absence of object cues in the environment. *Memory & cognition*, *35*, 1841–1853. doi: 10.3758/bf03192919

Steffens, M. C., Jelenec, P., & Mecklenbräuker, S. (2009). Decomposing the memory processes contributing to enactment effects by multinomial modelling. *European Journal of Cognitive Psychology*, *21*, 61–83. doi: 10.1080/09541440701868668

Steffens, M. C., von Stülpnagel, R., & Schult, J. C. (2015). Memory Recall after “Learning by Doing” and “Learning by Viewing”: Boundary Conditions of an Enactment Benefit. *Frontiers in Psychology*, *6*. doi: 10.3389/fpsyg.2015.01907

Stevanoni, E., & Salmon, K. (2005). Giving Memory a Hand: Instructing Children to Gesture Enhances their Event Recall. *Journal of Nonverbal Behavior*, *29*, 217–233. doi: 10.1007/s10919-005-7721-y

Suengas, A. G., & Johnson, M. K. (1988). Qualitative effects of rehearsal on memories for perceived and imagined complex events. *Journal of Experimental Psychology: General*, *117*, 377–389. doi: 10.1037/0096-3445.117.4.377

Tenpenny, P. L., Keriazakos, M. S., Lew, G. S., & Phelan, T. P. (1998). In search of inadvertent plagiarism. *The American Journal of Psychology*, *111*, 529–559. doi: 10.2307/1423550

Thomas, A. K., & Bulevich, J. B. (2006). Effective cue utilization reduces memory errors in older adults. *Psychology and Aging*, *21*, 379–389. (00031) doi: 10.1037/0882-7974.21.2.379

Thomas, A. K., Bulevich, J. B., & Loftus, E. F. (2003). Exploring the role of repetition and sensory elaboration in the imagination inflation effect. *Memory & Cognition*, *31*, 630–640. doi: 10.3758/bf03196103

Troyer, A. K., & Craik, F. I. M. (2000). The effect of divided attention on memory for items and their context. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *54*, 161–171. doi: 10.1037/h0087338

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, *80*, 352–373. doi: 10.1037/h0020071

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the norms. *Journal of Memory and Language*, *50*, 289–335. doi: 10.1016/j.jml.2003.10.003

Vetter, P., & Wolpert, D. M. (2000). Context estimation for sensorimotor control. *Journal of Neurophysiology*, *84*, 1026–1034.

von Essen, J. D., & Nilsson, L.-G. (2003). Memory effects of motor activation in subject-performed tasks and sign language. *Psychonomic bulletin & review*, *10*, 445–449. doi: 10.3758/bf03196504

Von Stülpnagel, R., & Steffens, M. C. (2012). Can active navigation be as good as driving? A comparison of spatial memory in drivers and backseat drivers. *Journal of experimental psychology: applied*, *18*, 162–177. doi: 10.1037/a0027133

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & Iverson, G. J. (2004). Assessing model mimicry using the parametric bootstrap. *Journal of Mathematical Psychology*, *48*, 28–50. doi: 10.1016/j.jmp.2003.11.004

Watanabe, R., Higuchi, T., & Kikuchi, Y. (2013). Imitation behavior is sensitive to visual perspective of the model: an fMRI study. *Experimental Brain Research*, *228*, 161–171. doi: 10.1007/s00221-013-3548-7

Watkins, M. J., & Gardiner, J. M. (1979). An appreciation of generate-recognize theory of recall. *Journal of Verbal Learning and Verbal Behavior*, *18*, 687–704. doi: 10.1016/s0022-5371(79)90397-9

Wickens, T. D. (2002). *Elementary Signal Detection Theory*. Oxford University Press.

Wickens, T. D. (2011). Multidimensional models for item recognition and source identification. In A. S. Benjamin (Ed.), *Successful Remembering and Successful Forgetting: A Festschrift in Honor of Robert A. Bjork* (pp. 1–22). London: Psychology Press.

Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception*, *44*, 289–300. doi: 10.1068/p7908

Wohlschläger, A., Gattis, M., & Bekkering, H. (2003). Action generation and action perception in imitation: an instance of the ideomotor principle. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *358*, 501–515. doi: 10.1098/rstb.2002.1257

Wutte, M. G., Glasauer, S., Jahn, K., & Flanagin, V. L. (2012). Moving and being moved: Differences in cerebral activation during recollection of whole-body motion. *Behavioural Brain Research*, *227*, 21–29. doi: 10.1016/j.bbr.2011.09.042

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: a review. *Psychological Bulletin*, *133*, 800–832. (00238) doi: 10.1037/0033-2909.133.5.800

Zaragoza, M. S., & Lane, S. M. (1998). Processing resources and eyewitness suggestibility. *Legal and Criminological Psychology*, *3*, 305–320. doi: 10.1111/j.2044-8333.1998.tb00368.x

Zimmer, H. D., Cohen, R. L., Guynn, M. J., Engelkamp, J., Kormi-Nouri, R., & Foley, M. A. (Eds.). (2001). *Memory for Action. A Distinct Form of Episodic Memory?* Oxford: Oxford University Press.

Zimmer, H. D., & Engelkamp, J. (1985). An attempt to distinguish between kinematic and motor memory components. *Acta psychologica*, *58*, 81–106. doi: 10.1016/0001-6918(85)90036-8

Zimmer, H. D., & Engelkamp, J. (1999). Levels-of-processing effects in subject-performed tasks. *Memory & Cognition*, *27*, 907–914. doi: 10.3758/BF03198543

Zimmer, H. D., & Engelkamp, J. (2003). Signing enhances memory like performing actions. *Psychonomic Bulletin & Review*, *10*, 450–454. doi: 10.3758/BF03196505

Zwickel, J., Grosjean, M., & Prinz, W. (2007). Seeing while moving: Measuring the online influence of action on perception. *The Quarterly Journal of Experimental Psychology*, *60*, 1063–1071. doi: 10.1080/17470210701288722

Zwickel, J., Grosjean, M., & Prinz, W. (2010). What part of an action interferes with ongoing perception? *Acta psychologica*, *134*, 403–409. doi: 10.1016/j.actpsy.2010.04.003

Zwickel, J., & Prinz, W. (2012). Assimilation and contrast: The two sides of specific interference between action and perception. *Psychological research*, *76*, 171–182. doi: 10.1007/s00426-011-0338-3