1   Running head: ACTION PREDICTION BASED ON STATISTICAL LEARNING

2

3

4   Translating visual information into action predictions: Statistical

5   learning in action and non-action contexts

6

7   Claire D. Monroy[a, c], Sarah A. Gerson[a, b] and Sabine Hunnius[a]

8   [a]Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen

9   Montessorilaan 3, 6500HE Nijmegen, The Netherlands, s.hunnius@donders.ru.nl

10   [b]School of Psychology, Cardiff University, 70 Park Place, Cardiff, CF10 3AT,

11   GersonS@cardiff.ac.uk

12   [c]Department of Otolaryngology, Ohio State University Wexner Medical Center,

13   claire.monroy@osumc.edu

14

15

16   Correspondence concerning this article should be addressed to: Claire Monroy, Department of

17   Otolaryngology, Ohio State University Wexner Medical Center, Email: claire.monroy@osumc.edu

18

19

20

21

22

23

24
25
26
27
28   Abstract
29

30   Humans are sensitive to the statistical regularities in action sequences carried out by
31 others. In the current eye-tracking study, we investigated whether this sensitivity can support
32 prediction of upcoming actions when observing unfamiliar action sequences. In two between-
33 subjects conditions, we examined whether observers would be more sensitive to statistical
34 regularities in sequences performed by a human agent vs. self-propelled 'ghost' events.
35 Secondly, we investigated whether regularities are better learned when associated with
36 contingent effects. Both implicit and explicit measures of learning were compared between
37 agent and ghost conditions. Implicit learning was measured via predictive eye movements to
38 upcoming actions or events, and explicit learning was measured via uninstructed reproduction
39 of action sequences and verbal reports of the regularities. Findings revealed that participants,
40 regardless of condition, readily learned the regularities and made correct predictive eye
41 movements to upcoming events during online observation. However, different patterns in
42 explicit learning outcomes emerged following observation: participants were most likely to
43 recreate the sequence regularities and to verbally report them when they observed an actor
44 create a contingent effect. These results suggest that the shift from implicit predictions to
45 explicit knowledge of what has been learned is facilitated when observers perceive another
46 agent's actions and when these actions cause effects. Findings are discussed with respect to
47 the potential role of the motor system in modulating how statistical regularities are learned
48 and used to modify behavior.
49
50   Keywords: action prediction, action sequences, statistical learning, implicit and
51 explicit learning, eye-tracking
52
53 **1.0 Introduction**

54   Predicting the behavior of other people is central to social cognition and interaction.

55 As we observe others, we automatically predict the unfolding movements and future course

56 of their actions (Flanagan & Johansson, 2003). In everyday life, many of the actions we

57 observe are embedded within continuous, temporal sequences. Imagine the act of baking a

58 cake: this action is comprised of a continuous stream of individual action steps such as

59 gathering ingredients, measuring them into bowls, mixing things together, pouring batter into

60 a tin, and so forth. The ability to anticipate the upcoming events in a sequence is an indicator

61 that the observer possesses some knowledge of the overarching structure of the global action

62 and the relations between the individual steps. Perceiving the boundaries of the distinct

63 elements in a sequence and anticipating what follows is crucial for our cognitive system to

64 perceive the overarching activity as coherent and meaningful (Zacks & Tversky, 2001). In the

65 current study, we investigated whether statistical regularities in novel, unfamiliar sequences

66   support the ability to generate predictions of future events during observation[1]. Specifically,

67   we investigated whether observers make anticipatory gaze fixations to upcoming action

68   events based on their transitional probabilities alone, and whether they recreate learned

69   regularities in their own action performance following observation.

70   1.1 Statistical learning in the domain of action

71   Statistical learning (SL) refers to the ability to detect regularities from structured input

72   and operates across sensory domains (Conway & Christiansen, 2005; Krogh, Vlach, &

73   Johnson, 2013). From early in life, humans are sensitive to multiple sources of statistical

74   information in visual and auditory stimuli (Saffran, Johnson, Aslin, & Newport, 1999).

75   Converging evidence indicates that SL skills are rapid and automatic, often occurring without

76   the learner being consciously aware that he or she has learned anything at all (Turk-Browne,

77   Scholl, Chun, & Johnson, 2008). This has led to the assumption that SL is a domain-general

78   mechanism, with similar underlying computations and outcomes across sensory modalities.

79   However, there is also evidence that the outcomes of SL are specific to the modality in which

80   the stimuli are learned. For instance, one study (Conway & Christiansen, 2006) presented

81   participants with auditory, tactile, and visual sequences defined by respective artificial

82   grammars. Findings showed that sensitivity to statistical features was specific to each sensory

83   modality, suggesting that SL involves "distributed, modality-constrained subsystems"

84   (Conway & Christiansen, 2006; p.911).

85   Does sensitivity to statistical regularities extend to the domain of action? If so, does

86   SL operate in a domain-general manner across all forms of perceptual events, or are there

87   specialized subsystems that might facilitate SL particularly for observed *actions*? An initial

88   study on action sequence processing by Baldwin and colleagues (2008) demonstrated that

---

[1] Unlike the cake example above, the sequences used in the current study were abstract in the sense that they did not lead to a global action goal. This was to ensure that predictions could only be based on acquiring knowledge of the sequence regularities rather than prior knowledge about the overarching event structure.

89 observers can rely on statistical regularities to segment action streams into discrete steps,

90 even when transitional probabilities are the only information available for identifying action

91 segments. At a group level, participants' performance on this action segmentation task was

92 comparable with performance on similar tasks in the language domain. Developmental

93 research has demonstrated similar findings with preverbal infants (Roseberry, Richie, Hirsh-

94 Pasek, Golinkoff, & Shipley, 2011; Saylor, Baldwin, Baird, & LaBounty, 2007; Stahl,

95 Romberg, Roseberry, Golinkoff, & Hirsh-Pasek, 2014), showing that these segmentation

96 skills emerge early in development. Similarity in performance across studies has led

97 researchers to speculate that a common "statistical tracking mechanism" (Baldwin,

98 Andersson, Saffran, & Meyer, 2008, p. 1404) is shared between processing of action and

99 processing of other forms of perceptual stimuli.

100   Segmentation reveals whether observers demonstrate sensitivity to the sequence

101 structure after learning has occurred. Typical paradigms measure segmentation by the ability

102 to remember the items they had observed during a previous learning phase (e.g., Baldwin et

103 al., 2008; Saffran et al., 1997). However, current theories of action perception claim that

104 continual, automatic prediction of upcoming actions is a central feature of action processing

105 (Kilner, Friston, & Frith, 2007a, 2007b). Importantly, predicting the outcomes of ongoing

106 actions requires integrating prior knowledge about the most likely outcomes of the action

107 with incoming perceptual input. Though active motor experiences are one important source

108 of action knowledge (Calvo-Merino, Grèzes, Glaser, Passingham, & Haggard, 2006; Libertus

109 & Needham, 2010; Sommerville, Woodward, & Needham, 2005), motor experience alone is

110 insufficient to explain the full range of infant and adults' capabilities for learning about

111 actions (Hunnius & Bekkering, 2014). Statistical learning skills are therefore a candidate

112 mechanism for how humans learn and generate predictions about upcoming action steps

113 when observing novel, unfamiliar sequences (Ahlheim, Stadler, & Schubotz, 2014), though

114   direct evidence for this does not yet exist. As we discuss below, we hypothesized that

115   observing human action engages specialized cognitive processes that particularly facilitate

116   learning of observed action sequences, relative to visual event sequences.

117   1.2 Outcomes of learning: implicit and explicit measures

118   The outcomes of SL have long been a topic of debate; in particular, discussions focus

119   on whether and under what conditions SL results in explicit or implicit learning outcomes

120   (Perruchet & Pacton, 2006). Typical findings have shown that SL usually occurs

121   automatically and without conscious intent; people are often unaware of the regularities they

122   have learned (e.g., Haider et al., 2014; Turk-Browne, Jungé, & Scholl, 2005; Turk-Browne et

123   al., 2008). Behavioral indicators of implicit learning are typically revealed in faster reaction

124   times (Fiser & Aslin, 2002) or anticipatory eye movements (Marcus, Karatekin, &

125   Markiewicz, 2006) and participants are usually unaware of the subtle changes in their own

126   behavior as a result of learning. On the other hand, SL can also result in explicit knowledge

127   about what was learned (Bertels, Franco, & Destrebecqz, 2012; Esser & Haider, 2017b).

128   Explicit learning is typically measured by recognition or recall which requires "conscious, or

129   deliberate, access to memory for previous experiences" (Gomez, 1997, p. 166). In the current

130   study, we assessed multiple measures of learning to explore how the learned information is

131   transferred into behavior. If participants learned the statistical regularities, they could in

132   principle predict what would occur next and shift their gaze to the next event in the sequence.

133   If implicit knowledge from observation can be accessed and used to modify behavior,

134   participants could also reproduce the observed regularities and report knowledge about the

135   sequence structure.

136   1.3 The role of the motor system during action observation

137   Observing actions engages neural networks that differ from those involved in general

138   visual and attention processes (Adams, Shipp, & Friston, 2013; Ahlheim et al., 2014;

139 Schubotz & von Cramon, 2009). For instance, neuroimaging research has revealed the

140 existence of a network of sensorimotor brain regions, collectively termed 'action-observation

141 network' (AON), which are specifically engaged when observing another person's actions

142 (Gallese & Goldman, 1998; Kilner, 2011). Activity in the AON, also sometimes termed

143 *'motor resonance'* (Rizzolatti & Craighero, 2004) or 'simulation' (Blakemore & Decety,

144 2001), is thought to facilitate prediction of observed actions by simulating how one would

145 perform the action oneself. Predictive accounts of the motor system propose that we employ

146 our own motor system using an internal, feed-forward model to predict the behavior of other

147 people we observe (e.g., Kilner et al., 2007b).

148   In the context of embodied accounts of action observation, the motor system

149 facilitates efficient transformation of visual information into action knowledge in the

150 observer's motor system. Supporting evidence from a separate line of research on

151 observational learning shows that observers are consistently better at imitating and learning

152 novel tool functions when observing a human actor relative to any other form of visual

153 observation (for a review, see Hopper, 2010). These behavioral studies employed the use of a

154 so-called 'ghost display', a method in which objects appear to move on their own with no

155 agent intervention. In the current study, we adopted the ghost-display method to test the

156 hypothesis that the learning advantage when observing another human, relative to a non-

157 agent ghost display, extends to action predictions based on statistical learning.

158 1.4 The role of effects in continuous action sequences

159   Goal-directed actions typically result in perceivable effects, such as the sound of a

160 whistle as it is blown. Through repeated observation, these effects become linked to the

161 actions that consistently precede them and create 'bidirectional action-effect associations'

162 (Elsner & Hommel, 2001). Prior research suggests that it is the effects of actions themselves

163 that people anticipate when planning their own movements (Hommel, 1996). In the field of

164   implicit learning research, action-effects have been shown to enhance implicit sequence

165   learning when participants own motor responses result in predictable action-effects (e.g.,

166   Haider, Eberhardt, Esser, & Rose, 2014). Recent work suggests that they may also be

167   particularly important for transferring learning from implicit into explicit awareness (Esser &

168   Haider, 2017a, 2017b). These findings demonstrate that action-effect associations likely play

169   a central role in establishing the contextual knowledge needed for making action predictions.

170   Though much of this work has investigated action-effects in sequence learning of motor

171   responses (e.g., using the standard *serial reaction time task*), there is also evidence to suggest

172   that action-effects also guide our predictions during observation alone (Paulus, van Dam,

173   Hunnius, Lindemann, & Bekkering, 2011).

174   How do sensory effects influence *observers'* sensitivity to statistical regularities when

175   they are embedded within continuous sequences, as is the case during daily real-life

176   perception? Based on ideomotor theory (James, 1890) and the related action-effect principle

177   (Hommel, 1996), observers should be better at learning action contingencies when they are

178   paired with an effect even when they do not produce the effects themselves. A matter that has

179   not received much attention, however, is the fact that non-action visual events also result in

180   sensory effects, such as a crashing wave. So far, we have defined effects as *action*-effects to

181   be consistent with prior research, but it is possible that sensory effects lead to similar

182   bidirectional associations in any form of perceptual sequence. In fact, another recent theory

183   (Schubotz, 2007) suggests that prediction of sensory effects occurs within our sensorimotor

184   system and can be generalized to any form of perceptual event, whether action or not. On the

185   other hand, as we described above, evidence for enhanced learning from observing action

186   suggests action-effects should be perceived and learned qualitatively differently than the

187   effects of non-action perceptual events. In the current study, we manipulated whether

188   statistical regularities were paired with an action-effect to investigate the importance of

189   observed effects for action predictions.

190   1.5 The current experiment

191   The central focus of this study was to investigate whether observers spontaneously

192   exploit statistical information in continuous action sequences to predict upcoming actions.

193   Our experiment included two manipulations in order to target two primary components of

194   action processing: (a) the role of observing an actors versus a ghost display (Agent and Ghost

195   conditions; between-subjects), and (b) the influence of action effects versus lack of effects

196   (Effect and No-effect pairs; within-subjects). These were assessed using an anticipatory

197   fixation eye-tracking paradigm during action observation, which has been established as a

198   measure of visual predictions (Hunnius & Bekkering, 2010). In addition, we examined the

199   link between predictive looking during observation and subsequent action production. For

200   this third aim, post-observation action performance and verbal reports were analyzed as

201   complementary measures of implicit and explicit learning.

202   **2.0 Method**

203   2.1 Participants

204   Fifty university students participated in this study (25 in each condition [Agent and

205   Ghost]; 43 females; $M = 20.07$ years, *range* = 18-25 years, $SD = 2.29$). Participants were

206   recruited via an online system for students at the university and were awarded course credit

207   for participation. Seven participants were excluded from analyses for not meeting the

208   inclusion requirements for total looking time (see *Analysis* section), resulting in 43

209   participants in the final sample (23 in the Agent condition and 20 in the Ghost condition).

210   2.2 Stimuli

211   Participants' eye movements were recorded with a Tobii T60 eye-tracker (Tobii,

212   Stockholm, Sweden) with a 17" monitor. Participants sat approximately 60cm away from the

213   screen. Stimuli were presented with Tobii ClearView AVI presentation software and sounds

214   were played through external speakers.

215   Participants observed a full-screen (1280x1024 pixels) film of a sequence involving a

216   multi-object device that afforded six unique manipulations and a central, star-shaped light

217   (Figure 1). To avoid confusion, we will subsequently refer to the individual object

218   manipulations in the sequence as 'events', as in one condition they were human actions and

219   in the other they were object movements. The movies were filmed with a Sony HandyCam

220   video camera and edited using Adobe Premiere Pro Cs5 software. The same device used

221   during filming was presented to participants before and after the observation phase.
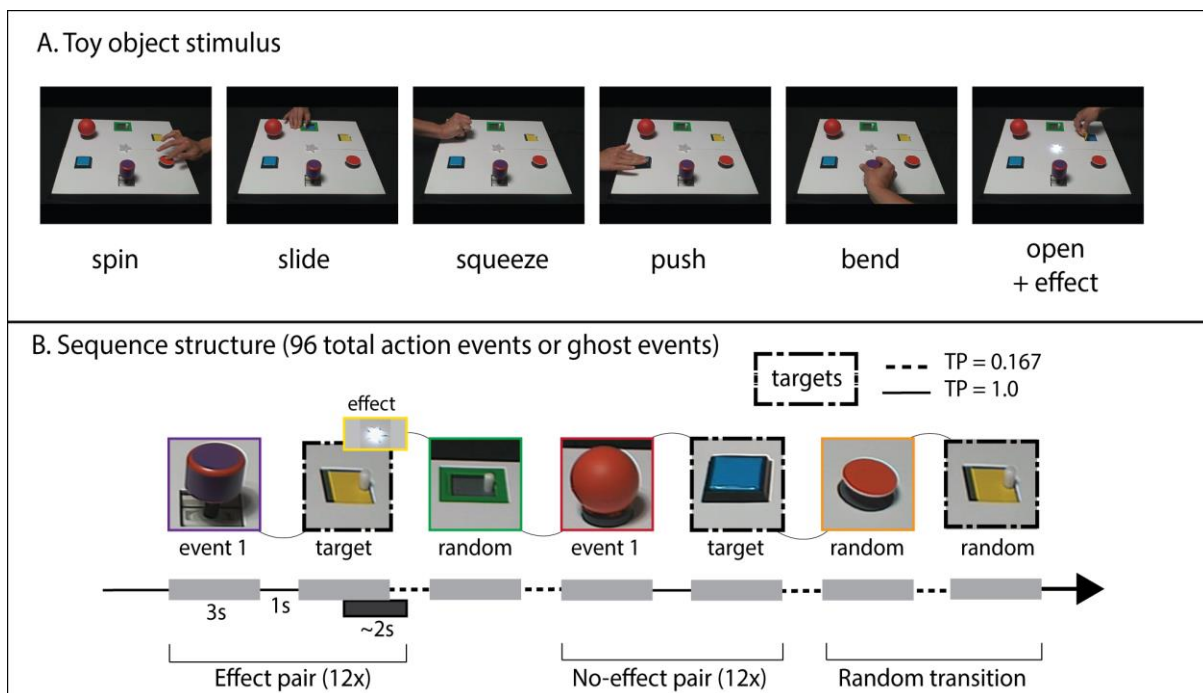
222
223   Figure 1. Overview of the experimental design.
224   A: Example frames from the video stimuli of the Agent condition. B: Schematic illustrating
225   the deterministic pairs and transitional probabilities within sequences during the observation
226   phase.
227

228   *2.2.1 Sequence*

229   We constructed four pseudo-randomized sequences, using the program *Mix* (van

230   Casteren & Davis, 2006). All sequences contained two deterministic pairs (transitional

231   probability between events = 1.0), labelled 'Effect' and 'No-effect' pairs (described in more

232 detail in the following paragraph). The second event of each deterministic pair was labelled a

233 *target*, as these were the events that became predictable as the sequence unfolded. All other

234 possible random pairs occurred with equal frequency (transitional probabilities between

235 events = 0.167; Figure 1B). No event or pair could occur more than three times consecutively.

236 All pairs and random events occurred 12 times (targets thus occurred 12 times within pairs

237 and 12 times outside of pairs). In total, participants viewed 24 deterministic pairs (12 Effect

238 and 12 No-effect pairs) and 48 random unpaired events, for sequences of 96 total actions or

239 events. Effect and No-effect pairs were composed of two actions that were randomly selected

240 from the 6 possible actions. Two sets of the four sequences were created: the two actions

241 comprising the Effect pair in one set became the No-effect pair in the second set, and vice-

242 versa. Thus, there were eight possible sequences within each condition and 16 videos in total;

243 participants were randomly assigned to view one of these videos.

244   The 'Effect pair' caused a central star to light up, whereas the 'No-effect pair' caused

245 no additional effect. We will subsequently refer to the second events of both pairs as *targets*,

246 as these were the events that became predictable as the sequence unfolded. The effect onset

247 occurred at a natural mid-point of the target event during the Effect pair: for example, during

248 the target *open*, the light turned on the moment the yellow door was fully open and turned off

249 again after it closed (see Figure 1A).

250   Targets could also occur elsewhere in the sequence outside of the deterministic pair

251 (see Figure 1B). In these instances, the effect never occurred. This ensured that the second

252 event did not independently predict the effect, and observers were required to learn the two-

253 step pair structure to accurately predict the effect.

254   Each video sequence was divided into four blocks, with the viewing angle oriented

255 from a different side of the box for each block. This was to dissociate the events (and their

256 corresponding objects) with their spatial location, and thus ensure that the observer could not

257   predict the next event based on its location on the screen. Each block lasted approximately

258   90s and consisted of 24 events. Brief cartoon animations were presented between blocks in

259   order to reengage the participant's attention. At the beginning of a block, one 4s still frame of

260   the stimulus was presented to allow observers to reorient to the new perspective. Movies

261   were approximately seven minutes long. Engaging, upbeat music was played throughout the

262   entire demonstration that did not correspond in any way to the unfolding sequence.

263   *2.2.2 Agent condition*

264   In the Agent condition movies, a hand manipulated the stimulus objects in a

265   continuous sequence. For each action, the hand entered the screen closest to the object on

266   which it acted. Each action was exactly three seconds in duration with a one-second pause

267   between actions during which the hand was off-screen and only the stimulus was visible.

268   *2.2.3 Ghost condition*

269   In the Ghost condition, the objects appeared to move on their own with a spotlight

270   focused on the current event (see Figure 2). The spotlight gradually illuminated each object

271   just prior to its movement onset and faded again after the object ceased moving. Between

272   ghost events, there was a 1s pause during which it was ambiguous where the spotlight would

273   next begin to appear, which matched the period of time the actor's hand was off-screen in the

274   Agent condition. Like the actor's hand, the spotlight cued which object would subsequently

275   move. The intensity and focus of the spotlight was equal for all objects. The sequence order

276   and timing of events were otherwise identical to the videos in the Agent condition.
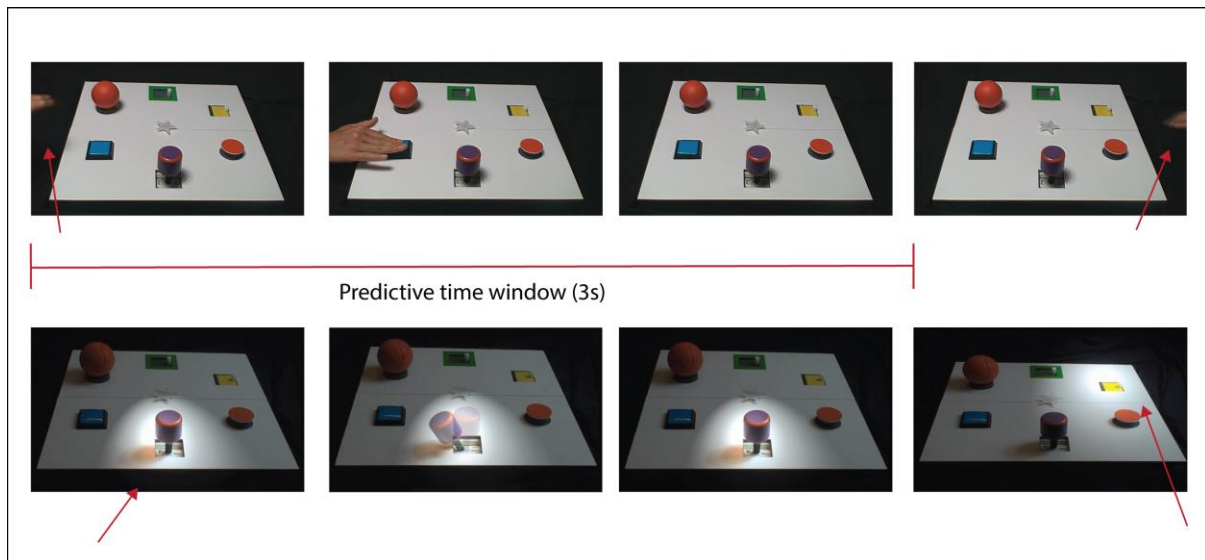
Figure 2. Predictive time windows in the Agent and Ghost conditions.
Example frames illustrating the predictive time windows in both conditions. Arrows indicate
the first frame in which the agent's hand appears (Agent condition) and in which the spotlight
focuses onto the target object (Ghost condition).

2.3 Procedure

Participants were first seated at a table upon which the stimulus device was placed.

The side facing each participant was counterbalanced. Participants were told they would

watch a video of a person interacting with the device, and were allowed to first familiarize

themselves with the objects before beginning the experiment. The side of the object facing

the participant during the action execution phase was kept the same as during the initial

familiarization. After familiarization, participants moved to a chair positioned in front of the

eye-tracking monitor for the observation phase in which they observed the stimuli videos.

First, the eye-tracker was calibrated using a standard 9-point calibration sequence provided

by Tobii Studio software. Calibration was repeated until valid calibration data was acquired

for at least eight calibration points. Following calibration, participants were shown one of the

eight stimulus sequences. They were told that they would be shown a video but were not

given specific viewing instructions.

296   Immediately after the observation phase, participants returned to the table and were

297   told that they could freely interact with the stimulus for one minute (this duration was based

298   on pilot testing). Participants were given no instruction, as our aim was to investigate whether

299   they would spontaneously integrate observed regularities into their own actions in the

300   absence of any task demand. The experimenter sat opposite the participant and monitored

301   their behavior, pressing a hidden button that activated the effect (i.e., central star light)

302   whenever he or she performed the Effect pair. After one minute, the experimenter ended the

303   action execution phase and then asked each participant the following questions: "Do you

304   know how to make the light turn on?" and "Did you notice any other pattern in the movies?"

305   If participants responded "yes" they were then asked to demonstrate the correct sequence on

306   the device. A camera facing the participant recorded this session and behavior was later

307   coded offline to assess action performance. Each participant completed one action sequence.

308   **3.0 Data Analysis**

309   3.1 Eye-tracking data

310   Participants with total fixation time more than one standard deviation below the mean

311   were excluded due to relative inattention to the movies. These participants yielded gaze data

312   for less than 25% of the demonstration, which corresponded to only 3 observations of each

313   pair and was insufficient to assess learning over the course of the experiment. This resulted in

314   the exclusion of two participants in the Agent condition and five participants in the Ghost

315   condition (see *Participants* section above).

316   Eye movement data was exported from Tobii ClearView analysis software and

317   separated into discrete fixations using a customized software program with a spatial filter of

318   30 pixels and a temporal filter of 100ms. Fixation data was imported into Matlab for further

319   analysis. Regions of interest (ROI) of identical size were defined around each object

320   (250x250 square pixels), and a smaller ROI (130x130 square pixels) was defined around the

321   light (due to its smaller size relative to the objects).

322   For the Agent condition, fixations were considered predictive if they occurred in the

323   time window from when the actor's hand entered the screen to perform the first action of a

324   pair until the frame before it reappeared for the target action (Figure 2). This corresponds to

325   the time in which the participant had enough information to predict the next action before its

326   onset. For the Ghost condition, this time window was defined from the moment the spotlight

327   highlighted the first object until the frame before the light shifted towards the second object

328   of a pair. Time windows were identical in length in both conditions. As the main aim of this

329   study was to examine prediction, only predictive gaze fixations were included in our analyses

330   (i.e., we did not examine reactive fixations).

331   To assess predictive gaze during observation, we compared proportions of fixations to

332   correct vs. incorrect objects (Implicit learning measure I). Implicit learning measure I reflects

333   the extent to which observers predict the correct location of an upcoming event, relative to

334   other locations. Second, we analyzed proportions of correct predictive fixations over the

335   course of the experiment to examine how learning unfolded over time. Third, proportions of

336   predictive fixations to target objects were compared between deterministic and random

337   transitions (Implicit learning measure II). Learning measure II reflects the frequency of

338   predictive looks to the target actions during predictable relative to non-predictable trials. We

339   describe both measures in more detail below (Figure 3).
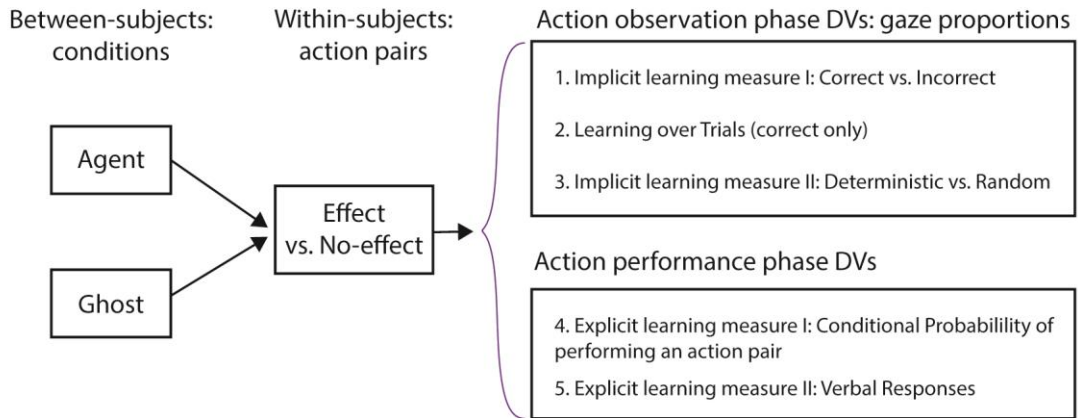
340

341   Figure 3. Overview of the experimental design and dependent variables.

342

343   *3.1.1 Implicit learning measure I: Correct vs. incorrect locations*

344   Target regions were defined around the location of the second events of each pair.

345   Fixations to targets during predictive time windows were counted as *Correct* and fixations to

346   the four remaining objects as *Incorrect*. Objects currently being manipulated (i.e., the first

347   action of the pair) were excluded from analyses. The first trial of each pair was not analyzed,

348   because participants were not expected to correctly predict the first observation of a pair. If

349   participants learned the pair structure, we expected them to make more fixations to the

350   locations of target objects relative to any other object during predictive time windows. For

351   both Effect and No-effect pairs, we calculated the proportion of correct and incorrect

352   fixations out of the total fixations to all objects (Eqs. 1 and 2). Because there were uneven

353   numbers of correct and incorrect locations, the incorrect proportion was defined as the

354   average number of fixations to the four remaining objects out of the total number of fixations.

355   This location measure represents observers' bias for looking toward the correct target,

356   relative to other objects, before it was acted upon. For additional analyses in which we

357   included fixations to the action-effect, see Supplementary Materials.

358
$$Correct_{target} = \frac{\# \, fixations \, to \, target}{total \, \# \, fixations \, to \, objects}$$
(1)

359
$$Incorrect_{target} = \frac{\#\ fixations\ to\ other\ 4\ objects/4}{total\ \#\ fixations\ to\ objects}$$
(2)

360   *3.1.2 Implicit learning measure II: Deterministic vs. random transitions*

361   Our second learning measure compared fixations to targets during deterministic vs.

362   random trials (Eqs. 3 and 4). Random trials were defined as transitions between any possible

363   event and the subsequent occurrence of a target event outside of a deterministic pair. We

364   discarded all repetition trials (for example, *push* followed by *push*) because it was impossible

365   to determine whether fixations during these trials were predictive or reactive (i.e., simply not

366   moving the eyes). This analysis thus enabled us to compare fixations to the same location

367   (target objects) in different statistical contexts.

368
$$Deterministic^2 = \frac{\#\ fixations\ to\ target\ (predictive\ trials)}{total\ \#\ fixations\ to\ objects}$$
(3)

369
$$Random = \frac{\#\ fixations\ to\ target\ (non-predictive\ trials)}{total\ \#\ fixations\ to\ objects}$$
(4)

370   3.2 Behavioral data

371   *3.2.1 Explicit learning measure I: Action performance*

372   Participants' self-produced action sequences were coded from the videotape

373   recordings. Each object manipulation was counted as a single action. We calculated the

374   conditional probability of performing the second action of a pair (*B*), given performance of

375   the first action (*A*), to account for variation in the overall length of participants' sequences.

376   Conditional probability was defined as:

377
$$P(B|A) = \frac{P(A,B)}{P(A)}$$
(5)

378   *3.2.2 Explicit learning measure II: Verbal responses*

379   Responses to the experimenters' explicit questions—"Do you know how to make the

380   light turn on?" and "Did you notice any other pattern in the movies?"—were coded as yes or

---

[2] Note that this equation is identical to Eq. 1

381   no; if their response was yes, it was further coded as 'yes'-correct or 'yes'-incorrect

382   depending on whether or not they demonstrated the correct sequence on the first attempt.

383   Proportions of participants who indicated each response type were calculated for each pair,

384   per condition.

385   **4.0 Results**

386   4.1 Eye movement data

387   To examine whether the Agent and Ghost displays elicited similar rates of overall

388   visual attention to the objects of interest, we compared the number of predictive fixations

389   between the two conditions. There were no differences in the number of anticipatory fixations

390   made during target trials (Ghost = 41.55, *SEM* = 4.80; Agent: *M* = 44.61, *SEM* = 3.41; *p* = .60)

391   or in the total number of fixations made across the entire demonstration (*p* = .21) suggesting

392   that differences in the visual stimuli in the Agent and Ghost conditions did not underlie any

393   potential differences in anticipatory fixations. Analyses of total looking times in seconds are

394   reported in the Supplementary materials.

395   *4.1.1 Implicit learning measure I: Correct vs. incorrect locations*

396   Our primary learning measures in each condition are presented in Table 1.

397   Proportions of gaze fixations were analyzed via a repeated-measure ANOVA with Prediction

398   (Correct vs. Incorrect) and Pair (Effect vs. No-effect) as within-subject factors and Condition

399   (Agent vs. Ghost) as a between-subjects factor. This analysis revealed a main effect of

400   Prediction, indicating that participants made a higher proportion of correct relative to

401   incorrect predictive fixations across pairs (mean difference = .14 [*SEM* = .04], *F*(1,40) =

402   16.27, *p* < .001, η_p^2 = .29). There were no other significant main effects or interactions

403   (*ps* > .13). The results of additional analyses including the location of the action-effect as a

404   correct location are available in the Supplemental Information.

405

406

407

408

409   Table 1.

410   *Main implicit and explicit dependent measures, separated by condition.*

Learning measure	Pair		Agent (*N* = 23)		Ghost (*N* = 20)	
			Mean	*SD*	Mean	*SD*
I: Correct vs. Incorrect	Effect	Correct (Eq. 1)	0.39	*0.26*	0.34	*0.33*
		Incorrect (Eq. 2)	0.09	*0.05*	0.11	*0.07*
	No-effect	Correct (Eq. 1)	0.26	*0.28*	0.25	*0.22*
		Incorrect (Eq. 2)	0.19	*0.07*	0.19	*0.06*
II: Deterministic vs. Random	Effect	Deterministic (Eq. 3)	0.39	*0.26*	0.34	*0.33*
		Random (Eq. 4)	0.25	*0.20*	0.18	*0.16*
	No-effect	Deterministic (Eq. 3)	0.26	*0.28*	0.25	*0.22*
		Random (Eq. 4)	0.13	*0.14*	0.14	*0.12*
Action Performance	Effect	Conditional probability (Eq. 5)	0.54	*0.36*	0.30	*0.30*
	No-effect		0.29	*0.36*	0.09	*0.16*
Verbal Response ("yes" – correct)	Effect	% participants	68.4%		15.4%	
	No-effect		5.9%		7.7%	

411   *Note*. For learning measure I, column 3 refers to proportions of correct and incorrect fixations. For learning
412   measure II, column 3 refers to proportions of correct fixations on deterministic or random trials.
413

414   *4.1.2 Learning over trials*

415   To examine changes in predictions across trials, we performed a general estimating

416   equations (GEE) analysis. GEE analyses are a preferred method for analyzing data with

417   repeated measures that contain missing points, such as trials in which no anticipatory

418   fixations were recorded, because they do not apply list-wise exclusion of cases (Zeger, Liang,

419   & Albert, 1988). Proportions of correct fixations to the targets were entered as the dependent

420   variable in a linear, model-based GEE with an unstructured Working Correlation Matrix.

421   Condition (between-subjects), Trial (within-subjects), and Pair (within-subjects) were entered

422   as predictors in a factorial model. In this analysis, the first trial was included (in contrast to

423   Learning measures I and II).

424   The GEE analysis yielded significant main effects of Trial ($\chi2(11) = 47.19$, $p < .001$)

425   and Pair ($\chi2(2) = 26.89$, $p < .001$) a significant interaction between Condition and Trial

426   ($\chi2(11) = 21.52$, $p = .028$) a significant interaction between Condition and Pair ($\chi2(2) = 8.70$,

427    $p = .003$) and a three-way Condition by Trial by Pair interaction ($\chi2(11) = 22.96$, $p = .02$).

428   The Condition by Pair interaction revealed that proportions of correct fixations were

429   significantly greater in the Agent relative to the Ghost condition for the Effect pair (mean

430   difference = .18 [$SEM = .05$], $p < .001$) but not for the No-effect pair (mean difference = -.09

431   [$SEM = .06$], $p = .11$)[3]. As illustrated in Figure 4, the Condition by Trial interaction revealed

432   that the Agent and Ghost conditions did not differ from one another on the very first ($p = .45$)

433   or second trial ($p = .15$). By the third trial, participants in the Agent condition made more

434   correct fixations than in the Ghost condition (mean difference = .28 [$SEM = .12$], p = .015)

435   and this pattern continued for several trials. The two conditions converged again by the 6th

436   trial ($p = .53$) for the remainder of the experiment. Together, these findings suggest that

437   participants showed a selective learning benefit for making correct anticipations when

438   viewing an agent producing action-effects, relative to the other observation contexts.
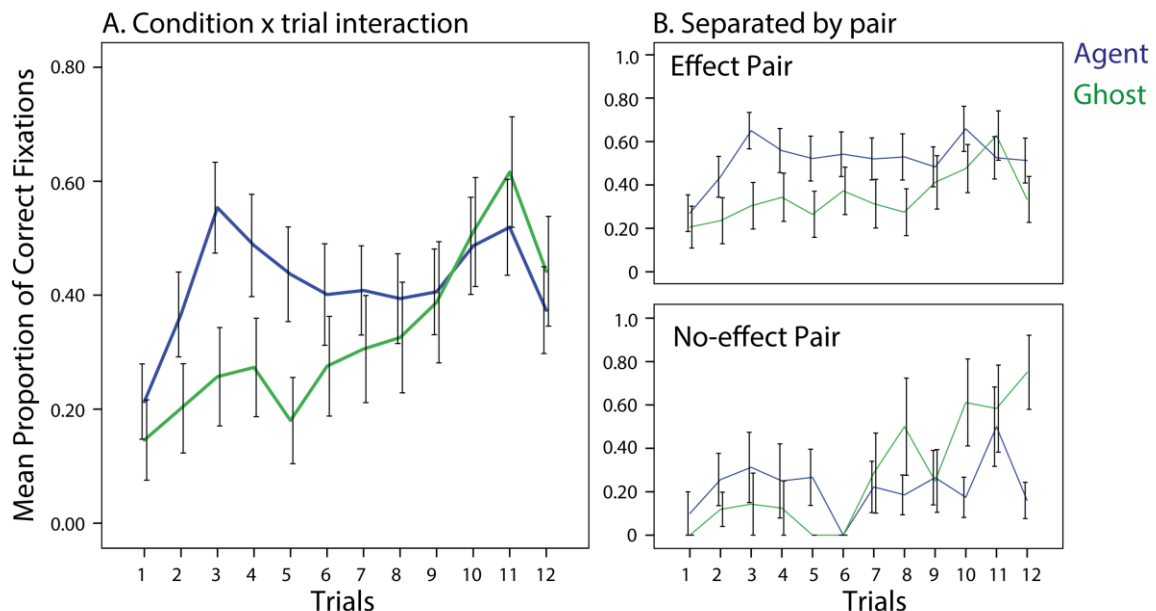
439

---

[3] Note that the interaction between Condition and Pair was not statistically significant in our first analysis (4.1.1). This is likely due to the fact that the first analysis included both correct and incorrect fixations, whereas the *Learning over Trials* analysis examines correct fixations only.

440 Figure 4. Learning over time.
441 Estimated marginal means of correct predictive fixations across pairs as a function of trial,
442 (left) collapsed across pairs and (right) separated by Effect and No-effect pairs. Bars
443 represent standard errors.
444

445   *4.1.3 Implicit learning measure II: Deterministic vs. random transitions*

446 The proportion of gaze fixations to target objects (Eqs. 3 and 4) were entered as the

447 dependent variables into an ANOVA with Transition (Deterministic vs. Random) and Pair

448 (Effect vs. No-effect) as within-subjects factors and Condition (Agent vs. Ghost) as a

449 between-subjects factor. This revealed a main effect of Transition, showing that participants

450 made more target fixations during deterministic than during random transitions across

451 conditions and pairs, $F(1, 42) = 42.9$, $p < .001$, $\eta_p^2 = .51$. There were no other effects or
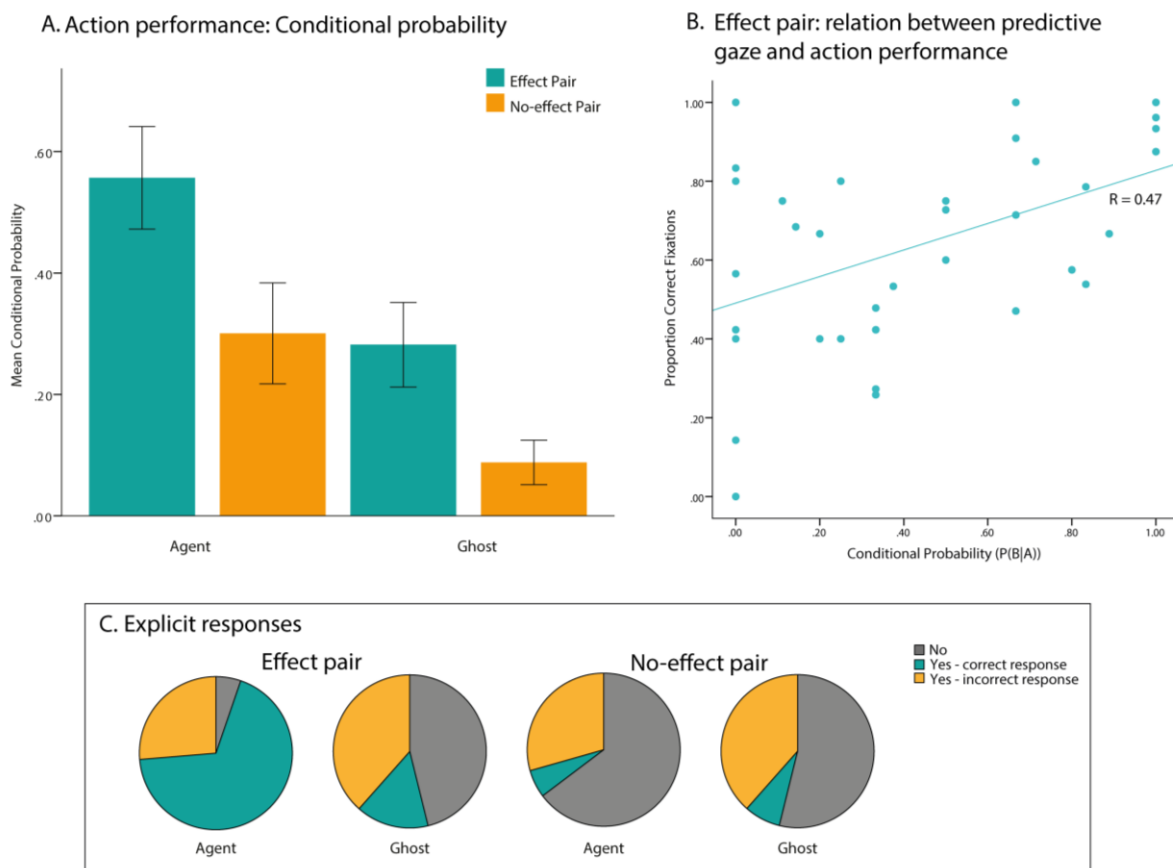
452 interactions ($ps > .11$).

453 4.2 Explicit measures of learning

454   *4.2.1 Explicit learning measure I: Action performance*

455   Across conditions, participants performed sequences with an average length of 26.22

456 actions ($SD = 7.1$), and performed a mean of 2.12 Effect pairs and 0.64 No-effect pairs (see

457 Table 1 for additional descriptive measures). There were no differences in the total length of

458 action sequences performed between conditions ($p = .19$).

459   Conditional probabilities for performing the target action given the performance of

460 the first action of the pair were entered in an ANOVA with Pair (Effect vs. No-effect) as a

461 within-subjects factor and Condition (Agent vs. Ghost) as a between-subjects factor. This

462 revealed main effects of Condition and Pair: participants in the Agent condition were more

463 likely to perform an action pair than those in the Ghost condition, $F(1, 34) = 11.57$, $p = .002$,

464 $\eta_p^2 = .25$ (see Figure 5a). Across conditions, participants were more likely to perform the

465 Effect pair than the No-effect pair, $F(1, 34) = 8.25$, $p = .007$, $\eta_p^2 = .20$. There was no

466 interaction between Pair and Condition ($p = .78$).

467   To assess whether participants in each group performed more pairs than would be

468   expected by chance, we conducted a one-sample t-test to compare the mean conditional

469   probability of performing each pair against a chance level of 0.167 (one out of six possible

470   actions, given any previous action). This revealed that the participants in the Agent condition

471   performed Effect pairs significantly more than chance ($p < .001$), while participants in the

472   Ghost condition did not ($p = .13$). In neither condition were the No-effect pairs performed at

473   an above-chance level ($ps > .05$).

Figure 5. Action performance and verbal awareness.
A: The mean probability of performing Effect and No-effect pairs ($P(B|A)$). Bars represent
standard errors. B: Scatterplot illustrating the relation between predictive fixations (Eq. 1)
and action performance (Eq. 5) for the Effect pair, across conditions. C: Pie graphs showing
the percentage of participants who gave each response type to the experimenter's question.
For the Effect pair, this was "Do you know how the light turns on?" and for the No-effect
pair this was "Did you see any other pattern in the movies?"

483   To investigate whether action execution was related to anticipatory looking behavior,

484   we correlated the proportion of correct target fixations (Eq. 1) and the conditional probability

485   of producing action pairs for each pair type. Across conditions, there was a significant

486   positive correlation between target fixations during Effect pairs and the conditional

487   probability of producing Effect pairs, $r(35) = .41$, $p = .02$, indicating that participants who

488   demonstrated higher rates of learning during the observation phase were more likely to

489   reenact the action-effect during the subsequent behavioral session (Figure 5b). There was no

490   correlation for the No-effect pair, $r(36) = .01$, $p = .97$. These correlation coefficients differed

491   significantly from one another, $Z = 1.75$, $p = .04$[4].

492   *4.2.2 Explicit learning measure II: Verbal responses*

493   Figure 5c illustrates the distributions of participants per each explicit response type to

494   the experimenter's questions following the action execution phase, separated by pair and

495   condition. The pie charts reflect the following pattern: 94.7% of participants in the Agent

496   condition reported explicit knowledge of the Effect pair; of these, 72.2% were correct and

497   27.8% were incorrect. Only 53.8% reported explicit knowledge of the pair in the Ghost

498   condition; 28.6% of these were correct and 71.4% were incorrect. Further, only 40% reported

499   knowledge of the No-effect pair across conditions, and those who did were usually incorrect

500   (93.3% of these 40%).

501   To compare these proportions of participants (Agent vs. Ghost) to one another, we

502   calculated the confidence intervals of the difference between them (the difference between

503   proportions is statistically significant wherever the confidence interval excludes zero;

504   Newcombe, 1998; Wilson, 1927). Table 2 reports the confidence intervals for the differences

505   in proportions for each response type. For the Effect pair, the proportion of participants who

506   responded 'yes' and were correct was significantly greater in the Agent than the Ghost

---

[4] For thoroughness, we also averaged across pairs and correlated the fixation proportions with conditional probability for Agent and Ghost conditions separately. Across pairs, there were no significant correlations for either group, ps > .42. These correlation coefficients did not differ significantly from one another ($Z = .41$, $p = .34$).

507   condition. A higher proportion of participants in the Agent condition reported knowledge of

508   the Effect pair—and could demonstrate the correct sequence—than in the Ghost condition.

509   Likewise, significantly more participants in the Ghost condition reported *no* knowledge of the

510   Effect pair than in the Agent condition. For the No-effect pair, the pattern of responses was

511   similar across conditions. Thus, participants observing an actor were more likely to retain

512   precise knowledge they could verbalize about the pair structure, but only when the actor's

513   actions led to a causal effect. Participants observing ghost events were less likely to report

514   verbal knowledge, and when they did, their representations of the pair structure were more

515   likely to be inaccurate.

516   **Table 2**.

517   *Mean differences (and confidence intervals) between conditions (Agent – Ghost) in the*

518   *proportions of participants reporting each response type for Effect and No-effect pairs.*

Response	Effect Pair		No effect Pair	
	Diff(P_a-P_b)	95% CI	Diff(P_a-P_b)	95% CI
"No"	-.41	**[.11, .66]***	.11	[-.22, .41]
"Yes"-correct	.53	**[.18, .73]***	-.02	[-.20, .28]
"Yes"-incorrect	-.12	[-.18, .42]	-.09	[-.22, .40]

519 *Note.* Diff(P_a-P_b) indicates the difference between the proportions of participants in the Agent and Ghost
520 conditions. *denotes statistically significant difference between the two sample proportions ($p < .05$).
521
522 **5.0 Discussion**

523   The current study investigated whether observers can learn statistical regularities

524   during observation of continuous action or event sequences. Specifically, we measured

525   anticipatory gaze fixations as an implicit measure of whether participants could use statistical

526   information to predict upcoming actions or events in the sequence. After learning, we

527   measured spontaneous action performance and verbal reports as explicit measures of whether

528   observed statistical regularities influence participants' self-produced actions and knowledge

529   of the sequence.

530   5.1 Implicit learning: Predictive gaze

531   Across conditions and pairs, participants demonstrated a robust tendency to predict

532   correct relative to incorrect locations. They also predicted the target more frequently during

533   deterministic relative to random transitions between events. In other words, they looked to

534   where a target event was statistically likely to occur next, and they looked to the targets

535   selectively when they were likely to occur next relative to when they were unlikely to occur

536   next.

537   When examining correct predictions over time, an interaction effect between these

538   two manipulations emerged: participants appeared to learn the regularities best when they

539   observed an actor produce an action-effect. In addition, different patterns emerged between

540   the Agent and Ghost conditions for implicit and explicit learning outcomes, as measured by

541   visual anticipations, action performance, and verbal knowledge of the pair structure.

542   Specifically, observing actions in the Agent condition did not seem to uniquely benefit

543   predictive gaze performance relative to observing visual events in the Ghost condition;

544   however, it did increase reproduction of the action pair and verbal knowledge about the pair

545   structure. Importantly, these differences were apparent only for the sequence pair which

546   resulted in an action-effect. One explanation for these patterns is that action-specific

547   processing in the Agent condition facilitated transfer from implicit (i.e., eye movements) to

548   explicit (i.e., self-produced actions, verbal awareness) knowledge, as we discuss in the

549   following sections.

550   5.2 Actions versus perceptual sequences

551   Participants demonstrated learning both when observing an actor and ghost events, as

552   indicated by their correct predictive looks while observing the sequences in both conditions.

553   This finding suggests that statistical learning operates consistently across the different types

554   of perceptual events, both action and non-action. Interestingly, learning emerged earlier in the

555   Agent condition than in the Ghost condition. Consistent with prior research, this finding

556   reveals a subtle learning benefit when observing an agent relative to other forms of visual

557   displays (Hopper, Flynn, Wood, & Whiten, 2010; Hopper, Lambeth, Schapiro, & Whiten,

558   2015). According to motor-based accounts of action observation, this benefit originates from

559   internal predictive models based in the motor system (Kilner et al., 2007; Stapel, Hunnius,

560   Meyer, & Bekkering, 2016). Here we show that observers demonstrate faster learning in the

561   Agent condition relative to the Event condition. Specifically, participants' rates of correct

562   fixations to target actions increased more quickly in the Agent condition, revealing that they

563   more easily detected the statistical relations between the actions and could modify their

564   looking behavior accordingly. Interpreted within these motor-based accounts, this may reflect

565   a more efficient ability to transfer knowledge acquired from visual statistical learning into

566   action predictions that are generated in the motor system (Kilner, 2009).

567   As discussed in the introduction, developmental studies have shown that children

568   learn significantly better from observing an agent performing actions relative to other forms

569   of observational learning (Hopper, 2010). One recent study, in fact, showed that toddlers

570   were able to learn action sequences when observing an actor, but not ghost events (Monroy,

571   Gerson, & Hunnius, 2017). This finding may reflect an interesting developmental shift, in

572   which actions provide a unique context that helps infants and children use acquired

573   knowledge from statistical learning to make predictions, above and beyond other stimuli.

574   Adults, on the other hand, are able to employ their statistical learning abilities across action

575   and non-actions contexts. Nevertheless, observing actions seems to elicit a learning benefit

576   that is consistent across development.

577   Though we made every attempt to match the stimuli in the two conditions for saliency,

578   there could have still been perceptual differences between the Agent and Ghost conditions

579   that could alternatively explain our findings. However, perceptual differences cannot solely

580   explain the observed results, as we find no differences in overall visual attention or predictive

581   fixations between conditions during observation. Secondly, both conditions demonstrated

582   learning during observation, but those in the Agent condition specifically reproduced more

583   action pairs and acquired more explicit sequence knowledge than participants in the Ghost

584   condition. This finding suggests that there were qualitative differences in the way the

585   sequence information was learned in Agent condition that are unlikely to be a result of

586   perceptual saliency.

587   5.3 The role of effects

588   Observing an agent produce causal effects led to higher rates of verbal knowledge and

589   reproduction of the action pair, relative to observing the ghost events or the pairs with no

590   effect (both action and ghost). This pattern supports the interpretation that observing actions

591   primarily influences the way in which learned knowledge is subsequently used to modify

592   behavior. Even though participants were uninstructed, observing an actor produce an effect in

593   the world may have automatically induced participants to perceive these events as goal-

594   directed, and to attempt to re-create them in the test setting. An alternative explanation,

595   suggestive of lower-level accounts, is that the action-effect simply provides additional

596   information and is therefore easier to learn. The action-effect relation contains more

597   information (i.e., A predicts both B and C) than the action-only pair (A predicts B). In

598   addition, the action-effect contingency contains an additional dimension (i.e., actions and

599   effects versus only actions). According to the model of sequence learning given by Keele and

600   colleagues (2003), multidimensional learning requires additional attention components that

601   are not required during unidimensional learning. These attentional requirements enhance

602   sequence learning by making the learned information accessible to explicit awareness (Keele,

603   Ivry, Mayr, Hazeltine, & Heuer, 2003).

604   When only analyzing correct predictions over time, an interaction effect emerged

605   which revealed that participants in the Agent condition demonstrated more correct predictive

606   fixations for the Effect relative to the No-effect pair, whereas this pattern did not hold for the

607   participants in the Ghost condition. However, this interaction effect did not appear when

608   comparing fixations to both correct and incorrect locations. One possible explanation for this

609   inconsistency is that, in the absence of a visual effect, participants were free to engage in

610   more visual exploratory behaviors to the other objects, resulting in higher proportions of

611   incorrect fixations for the No-effect pair relative to the Effect pair.

612   5.4 Action performance and its relation to prediction

613   Across conditions, participants were more likely to reproduce the pair associated with

614   an effect than the pair without an effect. In addition, rates of performing the effect pair were

615   correlated with participants' predictive looking for this pair. Specifically, the more accurately

616   observers predicted the Effect pair, the more likely they were to reproduce the effect

617   following observation. Adults and children easily recreate effects that they see in the world

618   when explicitly asked to do so; this has been empirically demonstrated in both forced-choice

619   and free-choice designs for simple action-effect contingencies (Elsner, 2007; Elsner &

620   Hommel, 2001). Here, our results provide new evidence that observers could recreate action-

621   effects based only on learning transitional probabilities, and they did so in the absence of

622   instruction or any explicit task. These findings suggest that new action knowledge—acquired

623   via observational statistical learning—can be accessed and used for action control when the

624   learned actions are used for produced a desired effect or outcome.

625   In addition, participants in the Agent condition were more likely to reproduce action

626   pairs than participants in the Ghost condition. This was not due to a general difference in

627   activity between the two conditions, as they did not simply perform more actions overall.

628   Based on the idea that we naturally tend to perceive human behavior as goal-directed,

629   observers in the Agent condition may have automatically attributed meaning to the actor's

630   actions and were more motivated to imitate what they observed, especially when they

631   resulted in an effect (Hopper, 2010; Hopper et al., 2014). Alternatively, consistent with the

632   faster emergence of correct anticipations in the Agent condition, these participants may have

633   also been better able to retain the new knowledge gained from the observed sequence and

634   apply it when performing their own action sequences than those in the Ghost condition.

635   5.5 Relations between predictive gaze, action performance, and verbal knowledge

636   Whether statistical learning engages implicit or explicit processes—and whether the

637   resulting knowledge is also implicit or explicit—is an ongoing debate (see Daltrozzo &

638   Conway, 2014 for a review). In the current study, we measured predictive gaze, action

639   performance, and verbal responses as reflecting different learning outcomes. These behaviors

640   may also relate to varying levels of implicit and explicit knowledge of the learned structure.

641   Studies on SL typically demonstrate that the outcomes of learning, and thus the learning

642   processes, are manifested in implicit behaviors such as anticipatory gaze, if at all (Fiser &

643   Aslin, 2001; Perruchet & Pacton, 2006; Turk-Browne et al., 2008). Currently, there is a

644   divide between those who argue that SL is an implicit mechanism (e.g., Clegg, DiGirolamo,

645   & Keele, 1998) and those who suggest that the process may be implicit but the knowledge

646   obtained via SL can become explicit when, for instance, learning reaches a certain threshold

647   (Cleeremans, 2006). In the former case, it is argued that knowledge can only become explicit

648   when other cognitive systems come into play. Recent findings have shown that sequence

649   learning also results in explicit knowledge depending on the 'task set'; that is, the relation

650   between the stimulus characteristics and the required response of the learner (Esser & Haider,

651   2017a, 2017b).

652   Consistent with these recent findings, our data suggest that observing action

653   sequences results in both implicit and explicit learning outcomes[5]. One possibility, grounded

654   in predictive accounts of the motor system, is that the knowledge gained via statistical

655   learning can be accessed by the motor system and used to update internal action models.

656   These models serve to generate predictions about the most likely upcoming action and to

657   prepare appropriate motor responses. Our findings differ from prior research in that, in the

658   current experiment, no response was required from participants during observation. Thus, the

659   resulting explicit knowledge did not arise from learned stimulus-response associations (as in

660   Haider et al., 2014). Rather, observation alone was sufficient to elicit both implicit and

661   explicit knowledge. Further, our findings suggest that observing human actions facilitates

662   both implicit sequence learning (indicated by faster learning rates in the Agent condition) and

663   transferring learned knowledge into explicit responses. However, as suggested by Schubotz

664   (2007), motor-based learning and prediction can still occur for external events (i.e., non-

665   actions). A fascinating question for further research is whether observing action sequences

666   engages entirely distinct learning processes from other forms of observational learning, or

667   whether the difference mainly lies in how the knowledge is accessed and used. Another

668   possibility to be considered is that acting immediately prior to being questioned by the

669   experimenter may have influenced some participants' verbal knowledge. That is, action

670   performance may have helped them to verbalize knowledge that otherwise would have

671   remained implicit. However, if there was an effect of acting on participants' explicit

672   knowledge of the sequence, this should have been consistent across conditions. Instead, the

673   dramatic group differences in verbal knowledge that we observed suggest that responses were

---

[5]As we did not directly measure the learning processes, but rather the learning outcomes, we cannot speak to whether or not the learning processes themselves were implicit or explicit and focus our discussion on the outcomes of learning.

674   primarily influenced by the action observation condition, rather than their own action

675   production.

5.6 Conclusion

677   The current study investigated whether SL abilities can support online prediction

678   during action observation. In particular, we compared observers' sensitivity to statistical

679   regularities in action sequences when observing a human actor relative to visual events. Our

680   main finding revealed that implicit learning occurred in both observation conditions and was

681   not dependent on action-effects; however, explicit knowledge was only consistently extracted

682   when observers viewed a human actor perform action sequences with causal effects. These

683   findings shed light on the potential role of the motor system in enhancing how information

684   learned solely via observation can be accessed and used to modify behavior.

690

691

692

693

694

695

696

697

698

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5

Ahlheim, C., Stadler, W., & Schubotz, R. (2014). Dissociating dynamic probability and predictability in observed actions-an fMRI study. *Frontiers in Human Neuroscience*, *8*, 273. https://doi.org/10.3389/fnhum.2014.00273

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–1407. https://doi.org/10.1016/j.cognition.2007.07.005

Bertels, J., Franco, A., & Destrebecqz, A. (2012). How implicit is visual statistical learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1425–1431. https://doi.org/10.1037/a0027210

Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*(8), 561–567. https://doi.org/10.1038/35086023

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, *16*(19), 1905–1910. https://doi.org/10.1016/j.cub.2006.07.065

Cleeremans, A. (2006). Conscious and unconscious cognition: A graded, dynamic perspective. *International Journal of Psychology*, *1*, 401–418. https://doi.org/10.4324/9780203783122

Clegg, B. A., DiGirolamo, G. J., & Keele, S. W. (1998). Sequence learning. *Trends in Cognitive Sciences*, *2*(8), 275–281. https://doi.org/10.1016/S1364-6613(98)01202-9

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of

724   tactile, visual, and auditory sequences. *Journal of Experimental Psychology*, *31*(1), 24–

725   39. https://doi.org/10.1037/0278-7393.31.1.24

726   Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between

727   modalities: Pitting abstract against stimulus-specific representations. *Psychological*

728   *Science*, *17*(10), 905–912. https://doi.org/10.1111/j.1467-9280.2006.01801.x

729   Daltrozzo, J., & Conway, C. M. (2014). Neurocognitive mechanisms of statistical-sequential

730   learning: what do event-related potentials tell us? *Frontiers in Human Neuroscience*, *8*,

731   437. https://doi.org/10.3389/fnhum.2014.00437

732   Elsner, B. (2007). Infants' imitation of goal-directed actions: The role of movements and

733   action effects. *Acta Psychologica*, *124*(1), 44–59.

734   https://doi.org/10.1016/j.actpsy.2006.09.006

735   Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of*

736   *Experimental Psychology: Human Perception and Performance*, *27*(1), 229–240.

737   https://doi.org/doi.org/10.1037/0096-1523.27.1.229

738   Esser, S., & Haider, H. (2017a). Action–effects enhance explicit sequential learning.

739   *Psychological Research*, 1–17. https://doi.org/10.1007/s00426-017-0883-5

740   Esser, S., & Haider, H. (2017b). The emergence of explicit knowledge in a serial reaction

741   time task: The role of experienced fluency and strength of representation. *Frontiers in*

742   *Psychology*, *8*, 502. https://doi.org/10.3389/fpsyg.2017.00502

743   Fiser, J., & Aslin, R. (2001). Unsupervised statistical learning of higher-order spatial

744   structures from visual scenes. *Psychological Science*, *12*(6), 499–504.

745   Fiser, J., & Aslin, R. (2002). Statistical learning of higher-order temporal structure from

746   visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and*

747   *Cognition*, *28*(3), 458–467. https://doi.org/10.1037/0278-7393.28.3.458

748   Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*,

749   *424*(6950), 769–771. https://doi.org/10.1038/nature01861

750 Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory. *Trends in*

751   *Cognitive Sciences*, *2*(12), 493–501. https://doi.org/10.1016/S1364-6613(98)01262-5

752 Gomez, R. (1997). Transfer and complexity in artificial grammar learning. *Cognitive*

753   *Psychology*, *33*(2), 154–207. https://doi.org/10.1006/cogp.1997.0654

754 Haider, H., Eberhardt, K., Esser, S., & Rose, M. (2014). Implicit visual learning: How the

755   task set modulates learning by determining the stimulus-response binding.

756   *Consciousness and Cognition*, *26*(1), 145–161.

757   https://doi.org/10.1016/j.concog.2014.03.005

758 Hommel, B. (1996). The cognitive representation of action: Automatic integration of

759   perceived action effects. *Psychological Research*, *59*(3), 176–186.

760   https://doi.org/10.1007/BF00425832

761 Hopper, L. M. (2010). "Ghost" experiments and the dissection of social learning in humans

762   and animals. *Biological Reviews*. https://doi.org/10.1111/j.1469-185X.2010.00120.x

763 Hopper, L. M., Flynn, E. G., Wood, L., & Whiten, A. (2010). Observational learning of tool

764   use in children: Investigating cultural spread through diffusion chains and learning

765   mechanisms through ghost displays. *Journal of Experimental Child Psychology*, *106*(1),

766   82–97. https://doi.org/10.1016/j.jecp.2009.12.001

767 Hopper, L. M., Lambeth, S. P., Schapiro, S. J., & Whiten, A. (2015). The importance of

768   witnessed agency in chimpanzee social learning of tool use. *Behavioural Processes*, *112*,

769   120–129. https://doi.org/10.1016/j.beproc.2014.10.009

770 James, W. (1890). The Principles of Psychology. *Journal of the History of Philosophy*, *21*(2),

771   270–272. https://doi.org/10.1353/hph.1983.0040

772 Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural

773   architecture of sequence representation. *Psychological Review*, *110*(2), 316–339.

774   https://doi.org/10.1037/0033-295X.110.2.316

775   Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive*

776   *Sciences*, *15*(8), 352–357. https://doi.org/10.1016/j.tics.2011.06.005

777   Kilner, J. M., Friston, K. J., & Frith, C. D. (2007a). Predictive coding: An account of the

778   mirror neuron system. *Cognitive Processing*, *8*(3), 159–166.

779   https://doi.org/10.1007/s10339-007-0170-2

780   Kilner, J. M., Friston, K. J., & Frith, C. D. (2007b). The mirror-neuron system: a Bayesian

781   perspective. *Neuroreport*, *18*(6), 1–5. https://doi.org/10.1097/WNR.0b013e3281139ed0

782   Krogh, L., Vlach, H. A., & Johnson, S. P. (2013). Statistical learning across development:

783   Flexible yet constrained. *Frontiers in Psychology*, *3*, 598.

784   https://doi.org/10.3389/fpsyg.2012.00598

785   Libertus, K., & Needham, A. (2010). Teach to reach: The effects of active vs. passive

786   reaching experiences on action and perception. *Vision Research*, *50*(24), 2750–2757.

787   https://doi.org/10.1016/j.visres.2010.09.001

788   Marcus, D. J., Karatekin, C., & Markiewicz, S. (2006). Oculomotor evidence of sequence

789   learning on the serial reaction time task. *Memory & Cognition*, *34*(2), 420–432.

790   https://doi.org/10.3758/BF03193419

791   Newcombe, R. G. (1998). Interval estimation for the difference between independent

792   proportions: Comparison of eleven methods. *Statistics in Medicine*, *17*(8), 873–890.

793   Paulus, M., van Dam, W., Hunnius, S., Lindemann, O., & Bekkering, H. (2011). Action-

794   effect binding by observational learning. *Psychonomic Bulletin & Review*, *18*(5), 1022–

795   1028. https://doi.org/10.3758/s13423-011-0136-3

796   Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One

797   phenomenon, two approaches. *Trends in Cognitive Science*, *10*(5), 233–238.

798   https://doi.org/10.1016/j.tics.2006.03.006

799   Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of*

800   *Neuroscience*, *27*(1), 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

801   Roseberry, S., Richie, R., Hirsh-Pasek, K., Golinkoff, R. M., & Shipley, T. F. (2011). Babies

802   catch a break: 7- to 9-month-olds track statistical probabilities in continuous dynamic

803   events. *Psychological Science*, *22*(11), 1422–1424.

804   https://doi.org/10.1177/0956797611422074

805   Saffran, J., Johnson, E., Aslin, R., & Newport, E. (1999). Statistical learning of tone

806   sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

807   https://doi.org/10.1016/S0010-0277(98)00075-4

808   Saffran, J. R., Newport, E. L., Aslin, R., Tunick, R. a., & Barrueco, S. (1997). Incidental

809   language learning: Listening (and learning) out of the corner of your ear. *Psychological*

810   *Science*, *8*(2), 101–105. https://doi.org/10.1111/j.1467-9280.1997.tb00690.x

811   Saylor, M., Baldwin, D., Baird, J., & LaBounty, J. (2007). Infants' on-line segmentation of

812   dynamic human action. *Journal of Cognition and Development*, *8*(1), 113–128.

813   https://doi.org/10.1080/15248370709336996

814   Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new

815   framework. *Trends in Cognitive Sciences*, *11*(5), 211–218.

816   https://doi.org/10.1016/j.tics.2007.02.006

817   Schubotz, R. I., & von Cramon, D. Y. (2009). The case of pretense: Observing actions and

818   inferring goals. *Journal of Cognitive Neuroscience*, *21*(4), 642–653.

819   https://doi.org/10.1162/jocn.2009.21049

820   Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-

821   month-old infants' perception of others' actions. *Cognition*, *96*(1).

822   https://doi.org/10.1016/j.cognition.2004.07.004

823   Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., & Hirsh-Pasek, K. (2014).

824   Infants segment continuous events using transitional probabilities. *Child Development*,

825   *85*(5), 1821–1826. https://doi.org/10.1111/cdev.12247

826 Stapel, J. C., Hunnius, S., Meyer, M., & Bekkering, H. (2016). Motor system contribution to

827   action prediction: Temporal accuracy depends on motor experience. *Cognition*, *148*, 71–

828   78. https://doi.org/10.1016/j.cognition.2015.12.007

829 Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical

830   learning. *Journal of Experimental Psychology*, *134*(4), 552–64.

831   https://doi.org/10.1037/0096-3445.134.4.552

832 Turk-browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2008). Neural evidence

833   of statistical learning: Efficient detection of visual regularities without awareness.

834   *Journal of Cognitive Neuroscience*, 1934–1945.

835   https://doi.org/10.1162/jocn.2009.21131

836 van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization.

837   *Behavior Research Methods*, *38*(4), 584–589. https://doi.org/10.3758/BF03193889

838 Zacks, J., & Tversky, B. (2001). Event structure in perception and conception. *Psychological*

839   *Bulletin*, *127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3

840 Zeger, S. L., Liang, K. Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized

841   estimating equation approach. *Biometrics*, *44*(4), 1049–1060.

842   https://doi.org/10.2307/2531734

843 Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference.

844   *Journal of the American Statistical Association, 22,* 209-212.

845

846