3

4

5

6

7   Extending the sense of agency: associative mechanisms underlying experiences of

8   voluntary and involuntary actions

9

10

11   Short title: Extending sense of agency

12

13

14   Nima Khalighinejad and Patrick Haggard

15   Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, UK

16

17   Correspondence should be addressed to Patrick Haggard**,** Institute of Cognitive

18   Neuroscience, University College London, Alexandra House, 17 Queen Square, London

19   WC1N 3AR, UK. Tel: +44 (0)20 7679 1149, E-mail: p.haggard@ucl.ac.uk

20

21

24

25

26

27

28

29 **Abstract**

30 Sense of agency refers to the experience that links one's voluntary actions to their external

31 outcomes. It remains unclear whether this ubiquitous experience is *hardwired*, arising from

32 specific signals within the brain's motor systems, or rather depends on associative learning,

33 through repeated co-occurrence of voluntary movements and their outcomes. To distinguish

34 these two models, we asked participants to trigger a tone by a voluntary keypress action.

35 The voluntary action was always associated with an involuntary movement of the other

36 hand. We then tested whether the combination of the involuntary movement and tone alone

37 might now suffice to produce a sense of agency, even when the voluntary action was

38 omitted. Sense of agency was measured using an implicit marker based on time perception,

39 namely a shift in the perceived time of the outcome towards the action that caused it. Across

40 two experiments, repeatedly pairing an involuntary movement with a voluntary action

41 induced key temporal features of agency, with the outcome now perceived as shifted

42 towards the involuntary movement. This shift required involuntary movements to have been

43 previously associated with voluntary actions. We show that some key aspects of agency

44 may be transferred from voluntary actions to involuntary movements. An internal volitional

45 signal is required for the primary acquisition of agency, but, with repeated association, the

46 involuntary movement in itself comes to produce some key temporal features of agency over

47 the subsequent outcome. This finding may explain how humans can develop an enduring

48 sense of agency in non-natural cases like brain-machine interfaces.

49

50

51

52

53

54 **Significance Statement**

55 In everyday life, people feel in control of their voluntary actions, and their outcomes. This

56 'sense of agency' could reflect hard-wired brain signals for volition, or could be acquired by

57 repeated association between a goal-directed action and another event. By pairing voluntary

58 actions of one hand with involuntary movements of the other hand, we showed that key

59 aspects of agency experience can transfer from voluntary to involuntary movements. Our

60 results explain why one can feel fully in control of one's actions even when they are

61 performed automatically, without focal conscious attention. We suggest that sense of agency

62 depends on a metacognitive signal that is relatively non-specific. Our findings could guide

63 acquisition of voluntary control using neuroprosthetics and brain-machine interfaces.

64

65
66
67
68
69

70

71

72

73

74

75

76

77

78   \body

79   **Introduction**

80   In a series of brilliant experiments, Roger Sperry switched the nerves for flexion of the rat

81   hind leg with the nerves for extension. After that, whenever the bottom of the foot was

82   injured, the rat extended the foot instead of flexing it. Rats never learned to lift up the paw,

83   and "no adaptive functioning of the nervous system took place" (1). When the optic nerves

84   of salamanders were cut, and the eyeball rotated 180 degrees, salamanders saw upside

85   down for the rest of their lives (2). These experiments suggested that key sensorimotor brain

86   circuits are largely hardwired, and impervious to modification by experience.

87   "Sense of agency" refers to the capacity to control one's actions, and, through them, the

88   external world. Sense of agency is fundamental to instrumental and goal-directed actions,

89   and forms the cornerstone of humans' astonishing capacity to change their physical and

90   social environment (3). However, it remains unclear how the brain produces this distinctive

91   and important subjective experience. Some recent results have linked the sense of agency

92   to specific preparatory volitional signals in frontal (4) and/or parietal (5) areas, which then

93   trigger voluntary motor commands passing through the "final common path" (6) of the

94   primary motor cortex. Importantly, these 'volitional signals' were generated well before the

95   occurrence of both action and outcome, and were strongly correlated with the subjective

96   intention to move. Such theories suggest a hard-wired, Sperry-esque account of human

97   volition.

98   In contrast, associative theories of agency deny the special status of internal volitional

99   signals, and focus instead on co-occurrence of actions and outcomes. For example, in

100   *ideomotor theories,* repeated association of actions and outcomes means that, over time,

101   actions come to be represented primarily in terms of their anticipated outcomes or goal-

102   states. By the same association, activation of the neural code for the goal event is then able

103   to generate the voluntary action (7). Stronger versions of this view suggest that people

4

merely *infer* their own agency based on observing the combination of action and outcome. There is no direct mental access to the internal processes that cause our actions, and the experiences of will and agency are mere inferences, or even illusions (8).

Current computational models of motor control, such as the comparator model (9, 10) have also been used to explain sense of agency. During action execution, efferent signals from motor areas are compared to predictions about the sensory consequences of the actions, such as feedback from a moving limb, or from some other external outcome of the action. These, contain elements of both the hard-wired and the associative frameworks. On the one hand, the sense of agency could depend necessarily on hard-wired efferent motor signal. On the other hand, the predictions generated by this signal must be based on learning an internal model from previous associations between efferent signals and their sensory consequences, consistent with the associative framework. It remains unclear to what extent human sense of agency is based on such hardwired signals or on learned associations.

These models make different predictions about the possibility of transferring agency from a voluntary action to another, co-occurring event. Mental properties commonly transfer from representation of one event to representations of another, notably in classical conditioning (11), but it remains unclear whether this occurs also for sense of agency. We asked participants to trigger a tone by making a voluntary keypress action with one hand. The voluntary action was always associated with an involuntary movement of the other hand. We then tested whether the combination of involuntary movement and tone alone might suffice to produce a sense of agency over the tone, even when no voluntary action was now present. Theories based on hard-wired efferent signals predict no sense of agency in this condition, since the putative internal volitional signal for one's own voluntary actions is, by definition, absent for involuntary movements. In contrast, associative learning theories predict that repeated co-occurrence of a voluntary and an involuntary movement could produce associative transfer, so that involuntary movements could, by association, come to acquire the same sense of agency that characterises voluntary movements.

131   We therefore designed two between-subject experiments. 36 participants were recruited for

132   the first experiment and were randomly assigned to the experimental and control groups

133   (Fig.1). In the experimental group, self-paced voluntary keypress actions of the right hand

134   triggered an immediate and physically-similar involuntary keypress movement of the left

135   hand, imposed by a robotic arm (Phantom Premium, 3D Systems, South Carolina, USA).

136   These movements were followed by a tone 250 ms later, in the operant condition.

137   Participants could thus learn to associate voluntary action, involuntary movement, and tone.

138   Such "learning" trials alternated with "test" trials containing only involuntary movements

139   followed by tones, and no voluntary action. Sense of agency over the tone was measured

140   using an implicit marker based on time perception. Participants judged the time of the tone

141   using a rotating clock display. A shift in the perceived time of the tone towards the preceding

142   action, is an established implicit marker of agency. This shift is compared to a baseline

143   condition containing only a tone, but no action. Importantly, involuntary movements are not

144   sufficient to cause perceptual shifts of the tone, and a volitional signal appears necessary

145   (12–14). A further control group of participants also judged the time of the tone following an

146   involuntary movement, but had never experienced any association between involuntary and

147   voluntary movement. Kinematics of both hands' movements were monitored online using

148   accelerometers.

149   Experiment 2 used the same design, but triggered involuntary movements by non-invasive

150   brain stimulation. 36 participants were recruited and were randomly assigned to

151   experimental and control groups. Self-paced voluntary actions of one hand were now paired

152   with involuntary twitches of the other hand, caused by transcranial magnetic stimulation

153   (TMS) over primary motor cortex. These learning trials again alternated with test trials

154   containing only involuntary TMS-evoked twitches followed by tones. Motor evoked potentials

155   (MEPs) were recorded from the first dorsal interosseous of the left hand. The control group

156   also judged the time of the tone following an involuntary twitch, but they had never

157   experienced any association between the twitch and any voluntary actions.

158   If sense of agency depends on a hard-wired efferent signal from the voluntary motor system,

159   no amount of associative learning should be able to induce key temporal features of

160   experience of agency for involuntary movements followed by tones, because the necessary

161   volitional signal is absent in this case. Conversely, if sense of agency is based on

162   associative learning, and if such associations can transfer from volitional signals to other

163   events, the repeated association between voluntary action and involuntary movement should

164   suffice to support some key temporal features of experience of agency over a tone triggered

165   by involuntary movement.

166   **Results**

167   **Experiment 1. Involuntary movement induced by a robotic arm.** 36 participants were

168   randomly assigned to the experimental group (n=18) or the control group (n=18). Data from

169   four participants were lost due to technical errors, leaving 16 participants in each group. We

170   already knew, from previous evidence, that the perceived time of a tone shifts towards a

171   preceding voluntary action, but not an involuntary movement (12). Surprisingly, in the

172   experimental group, we also found a perceptual shift in the perceived time of the tone

173   towards the involuntary movements on interleaved test trials occurring in-between truly

174   voluntary actions (one-sample, $t(15)=-4.18$, $p<0.01$, 95% CI [-200, -65]). In the control group,

175   who never experienced association between voluntary actions and involuntary movements,

176   the perceived time of the beep did not shift towards the preceding movement (one-sample,

177    $t(15)=-1.46$, $p=0.17$, 95% CI [-89, 17]). Crucially, the tone binding was significantly stronger

178   in the test trials of the experimental group compared to the control group ($t(30) = -2.40$,

179    $p=0.02$, $d=0.85$, 95% CI [-179, -14]) (Fig.2, Fig.S1, Fig.S2 and Table S1).

180   Finally, to confirm that the difference in baseline blocks did not drive the observed effect,

181   judgement errors from the baseline blocks were compared. No significant difference was

182   observed between the experimental and the control group ($t(30) = -0.03$, $p=0.97$, $d=0.01$,

183   95% CI [-44, 43]).

Experiment 2. Involuntary movement induced by TMS. Experiment 2 aimed to replicate the first experiment and to explore any potential differences in the central and peripheral routs of passive movement induction. 36 participants were randomly assigned to the experimental group (n=18) or to the control group (n=18). Data from two participants were lost due to technical errors, leaving 17 participants per group. We replicated the core findings of experiment 1. In the experimental group, associating passive movements with voluntary actions of the other hand led to the perceptual shift of outcomes towards TMS-induced passive movements (one-sample, $t(16)=-3.27$, $p<0.01$, 95% CI [-133, -28]). No tone binding was observed in the control group (one-sample, $t(16)=-0.31$, $p=0.76$, 95% CI [-67, 50]). Direct comparison of the two groups showed a clear trend for stronger binding on test trials in the experimental group compared to the control group, though with a lower effect size that in experiment 1 ($t(32) = -1.96$, $p=0.06$, $d=0.67$, 95% CI [-147, 3]) (Fig.2, Fig.S1, Fig.S2, and Table S2).

Finally, to confirm that the difference in baseline blocks did not drive the observed effect, judgement errors from the baseline blocks were compared. No significant difference was observed between the experimental and the control group ($t(32)=0.04$, $p=0.97$, $d=0.01$, 95% CI [-48, 50]).

Experiment 1&2. To investigate the generality of the effect across experiments, we performed a 2x2x2 ANOVA with the within subject factor of condition (baseline vs. operant), between subject factor of experiment (Exp1 vs. Exp2) and the between subject factor of group (experimental vs. control). The significant interaction between condition and group ($F(1,62)=9.54$, $p=0.003$, $\eta2=0.13$) recapitulated previous findings. Post-hoc analysis showed that the difference in the perceived time of the beep between two groups lays in the operant condition ($t(64)=-2.54$, $p=0.01$, $d=0.63$, 95% CI [-150, -18]), not the baseline ($t(64)=0.01$, $p=0.99$, $d<0.01$, 95% CI [-32, 32]). There was no significant main effect of experiment ($F(1,62)=2.67$, $p=0.11$, $\eta2=0.04$). Importantly, there was no interaction between the condition and experiment ($F(1,62)=2.12$, $p=0.15$, $\eta2=0.03$), group and experiment ($F(1,62)=0.10$,

211 p=0.75, η2<0.01) or condition, experiment and group ($F_{(1,62)}=0.20$, p=0.66, η2<0.01). This

212 suggests that the observed effect is a general phenomenon, regardless of the method used

213 to induce passive movement (Fig. 2 and Fig. S1).

Discussion

215 'Intentional binding' refers to the perceived compression of an interval between voluntary

216 actions and their sensory consequences. In particular, participants reliably perceive the

217 sensory consequences of their voluntary actions as happening earlier in time compared to a

218 baseline condition where the same event occurs without a voluntary action. Importantly,

219 involuntary movements were reported not to produce the same binding of tones observed

220 after voluntary actions, but in fact produced a temporal repulsion (12).

221 Here, we showed, in two separate experiments, that repeatedly pairing an involuntary

222 movement with a voluntary action can lead to intentional binding with respect to the

223 involuntary movement. We have used tone binding as an implicit marker of sense of agency.

224 Specifically, voluntary-involuntary pairing led to tone binding on involuntary movement test

225 trials where voluntary action was absent. This acquisition of key temporal features of agency

226 for involuntary movements did not occur in a control group who never made voluntary

227 actions. Alternative explanations based on increased attention in the voluntary action group

228 could not explain the pattern of results observed in our data (see supplementary results and

229 Fig. S4). Thus, voluntary actions are necessary for the emergence of a sense of agency.

230 However, once a voluntary signal is present, it can be mentally associated with other events,

231 and spread to produce the distinctive intentional binding feature of volition, but now with

232 respect to other movements. Our results therefore suggest that some key temporal features

233 of experience of agency can be transferred by association from truly voluntary actions, to

234 movements that are, in fact, involuntary, and purely passive.

A path model of agency acquisition

9

236   In our everyday life we perceive our voluntary actions as caused by our intention to produce

237   a specific outcome. These voluntary actions are often associated with two specific

238   experiences: The experience of volition reflects the initiation and control of the voluntary

239   action, and possibly a prediction of the outcome. The experience of agency, in contrast, is

240   based on attributing the actual outcome back to one's own triggering action (15) (Fig.3.A). In

241   our experiment, we reprogrammed the experiences surrounding voluntary action, by making

242   participants perform two movements at the same time, one voluntary and the other

243   involuntary (during *learning* trials). Thus, the intention to initiate the voluntary action was

244   associated with two movements, one located on each hand (Fig.3.B). Classical intentional

245   binding predicts that experience of agency arises when there is both a direct relation

246   between a movement and its outcome (path 2 in Fig.3.C), and also a direct relation between

247   the movement and the intention which precedes it (path 1 in Fig.3.C). The necessity of path

248   1 is clear from previous results (12, 14) showing that intentional binding does not occur for

249   involuntary movements.

250   In our experimental group, a further path (path 3 in Fig.3.C), similar to path 2, also exists

251   between the involuntary movement and the outcome. Data from our experimental group

252   shows that this path can generate some key temporal features of agency, such as intentional

253   binding. Importantly, comparison with the control group shows that functioning of path 3

254   strongly depends on its previous association with internal volitional signals (path 1). For the

255   control group, the involuntary movement was never paired with the voluntary action, and

256   involuntary movements never showed the key temporal linkage to outcomes. This finding

257   suggests that a single volitional signal can drive multiple action-outcome relations. As a

258   result, some key temporal features of agency can arise for movements that are merely

259   correlated with an intention, but not directly caused by it. This, in turn, suggests that the

260   relation between intention and sense of agency is not precisely-matched, and is not effector-

261   specific (Fig.3.D).

262   **Can we be mistaken regarding the facts of our own agency?**

263 Explicit measures of agency are subject to a number of cognitive biases, and are highly

264 sensitive to task demands. We therefore advisedly chose an implicit measure of sense of

265 agency, based on time perception. Synofzik et al. (2008) suggested that sense of agency

266 comprises two different levels: an implicit 'feeling of agency' and an explicit 'judgement of

267 agency' (16). Based on this view, the 'feeling' of agency is produced implicitly by low-level

268 sensorimotor signals. In the rare case that one must explicitly judge one's own agency, this

269 low-level feeling of agency provides the primary evidence for the judgement. However, social

270 contextual cues and other priors can bias such judgements. The intentional binding task

271 focusses on the non-conceptual feeling of agency. We did not obtain explicit judgements of

272 agency in this task. As a result, we cannot know whether the involuntary movements of the

273 experimental group came to feel like "I did that", but we imagine they did not.

274 In healthy adults, voluntary and involuntary movements generate quite different experiences

275 (17), and our brief training was unlikely to suppress this difference. Indeed, most systems of

276 law are based on a 'voluntary action condition', which rigidly assumes a distinct subjective

277 experience of voluntary action (18). In particular, selection and preparation of action in

278 frontal motor areas appears essential for a full experience of voluntary control (19).

279 Nevertheless, our results show that some key features of sense of agency can be

280 transferred from voluntary to involuntary movements, given appropriate associative learning.

281 The experience of the tone following involuntary movement acquired some temporal features

282 of agency, but this does not imply that participants would judge themselves the author of the

283 tone. Here, we have used implicit measures to show that one key feature of voluntary action,

284 namely the important *'goal-directed'* or *'ideomotor'* feature, by which the experience of action

285 leads to anticipation of outcomes, can transfer to involuntary movements. Interestingly,

286 patients with psychosis may have a deluded experience of their own actions. These

287 frequently involve false positives, such as reporting voluntary control over external events

288 unrelated to their own actions, such as changing traffic lights, or news events.

289   Our results can also be interpreted using an active inference framework (20). Here,

290   intentions are abstract predictions about likely outcomes, which are Bayes-optimally

291   combined with sensory evidence about outcomes when this becomes available. Intentional

292   binding has been modelled as a Bayes-optimal integration of action and outcome (21). Thus,

293   strong tone binding might arise because intentional actions provide a high-precision prior for

294   estimating outcomes. We found that pairing a second event, in this case an involuntary

295   movement, with a high-precision intentional prior results in that event having a similar

296   influence on outcome perception to the original intentional action. Thus, the structuring

297   effects of voluntary action on outcome perception may not reflect some unique experiential

298   quality specific to volition (though see ref. 22), but simply that intentional actions normally

299   serve as high-precision priors for their outcomes.

300   **Specificity of Internal volitional signals underlying agency acquisition**

301   We conclude that an internal volitional signal is required for the acquisition of sense of

302   agency. However, after repeated association, the volitional signal is not required for

303   subsequent expression of key temporal features of agency, such as intentional binding.

304   Moreover, the putative volitional signal is not highly specific with respect to *which* agency

305   relations are established. In our case, volitional signals controlling the right hand lead to

306   intentional binding for involuntary movements of the left hand. Thus, a range of

307   movement/outcome pairings may be enabled by co-occurrence with intention. Intentions do

308   structure subsequent subjective experience, but by means of a loose fit, rather than a tight

309   prediction about specific muscular movements. Previous studies suggested that the sense of

310   agency is highly temporally-specific, in that intentions, actions, and outcomes must follow a

311   predictable temporal sequence (19, 23). However, the *content* of intention, action and

312   outcome can be combined arbitrarily without compromising the experience of agency. Our

313   result suggests that volitional signals have the interesting property of high "latent

314   associability": they potentiate the development of *any* operant relation they co-occur with.

315   This is consistent with Skinner's demonstration that animals assume a causal connection

316 between an action and a reinforcing stimulus, even when the connection is in fact an

317 accidental correlation (24).

318 In our case, the path between volition and agency is not effector-specific, but effector-

319 independent. In particular, our design involved voluntary actions and involuntary movements

320 assigned to different hands. Our results thus suggest that the contribution of internal

321 volitional signals to sense of agency is bihemispheric, rather than hemisphere-specific.

322 Rodent studies showed that mice readily learn to control a robot when arbitrary motor cortex

323 activity is used to drive the robot dynamics. Learning such intentional neuroprosthetic skills

324 depends on corticostriatal plasticity (25). Our results likewise show that linking formation of

325 an intention to an outcome leads to formation of some key temporal features of agency,

326 even when the means that mediate between intention and outcome are artificial, and even

327 after the original volitional signal is dropped. These findings may explain how humans can

328 develop an enduring and successful feeling of agency in cases of non-natural movement like

329 brain-machine interfaces (26).

330 **Learning one's own agency**

331 Our experiment suggests that a conjunction of three conditions may be sufficient for sense of

332 agency. First, an internal volitional signal must be present to provide a general metacognitive

333 experience of intentional action. Second, some body movement must occur. Third, some

334 external outcome of the action must occur. We also showed that no specific linkage between

335 the metacognitive volitional signal and the body movement is necessary. In particular, the

336 volitional signal need not be present at the same time as the body movement, nor even

337 relate to the same effector. Thus, the internal volitional signal need not have a hardwired

338 connection to the motor output system in the manner suggested by Sperry. In our

339 experiment, it was sufficient that the volitional signal and the body movement had previously

340 been associated.

341 This pattern of results reflects two fundamental features of human voluntary action, which we

342 call *automaticity* and *flexibility*. *Automaticity* refers to the way that actions which initially

343 require focussed attention, such as driving a car, or cooking soufflés, become increasingly

344 fluent with repetition. The subjective experience of action also changes. The action becomes

345 less central in conscious experience, and instead provides a background 'buzz' of

346 awareness (16). However outcomes are still fully attributed to one's own agency. Our results

347 show a similar retention of key temporal features of experience of agency even when our

348 experimental design deliberately reduced and removed intentional control over the outcome.

349 Thus, our study can clarify a striking paradox of human action: namely, that one can feel fully

350 in control of a skilled action such as riding a bicycle, and have a clear sense of agency, yet

351 have only thin conscious experience of the action itself.

352 *Flexibility* refers to the ability of humans and animals to achieve control over goal states

353 using complex and varying means (27). This perhaps contributes to the astonishing human

354 proficiency in developing and using technology. Hebb's classical concept of motor

355 equivalence (28) suggests that cognitive systems are not generally concerned whether a

356 goal is achieved with one effector or with another – all movements that achieve the goal are

357 effectively equivalent.

358 This transfer of key temporal features of experience of agency from intentional actions to

359 other movements recalls the way that sense of agency emerges in human development.

360 Human infants appear to act randomly, with little intentionality and goal-directedness,

361 compared to healthy adults. During early experience, infants may gradually learn the precise

362 mapping between different intentions, the resulting body movements, and external

363 consequences. They thus eventually acquire the capacity to move a specific effector –

364 achieving control over their body, and thus over their environment. Our results show that the

365 capacity to form new intention-movement-outcome associations seems to remain and,

366 importantly, could be generalised to non-voluntary movements, even when intentional action

14

367 is no longer present. In this regard, it has been shown that younger children tend to confuse

368 intended with accidental outcomes (29–31).

369 Our experiments suggest that a hardwired internal volitional signal is required for the initial

370 *acquisition* and emergence of sense agency. Importantly, this hardwired signal appears to

371 be cognitive rather than motoric, since it is not linked to any specific output effector. At the

372 same time, associative mechanisms contribute strongly to the *expression* of sense of

373 agency. The presence of internal volitional signals during learning (path 1 in Fig.3.C) is

374 necessary for induction, though not expression of key temporal features of experience for

375 both direct voluntary action (path 2 in Fig.3.C), and also for an associated involuntary

376 movements (path 3 in Fig.3.C).

377 Wittgenstein (17) famously asked "What is left over if I subtract the fact that my arm goes up

378 from the fact that I raise my arm?". Sense of agency is a partial answer to this question.

379 However, even simple voluntary actions trigger widespread and automatic involuntary

380 elements. For example, voluntarily lifting the right arm requires anticipatory compensations

381 in contralateral muscles (32, 33). Thus, voluntarily moving one effector normally leads to

382 involuntary (or at least less voluntary) adjustments elsewhere, rather as in our experimental

383 group. Importantly, people are not generally surprised, or even aware of these involuntary

384 adjustments – although they would presumably be immediately conscious of a comparable

385 passive displacement of the same body parts. Thus, the involuntary side-effects of voluntary

386 action come to form part of an integrated experience of agency (16). The highly distributed,

387 integrated nature of motor control ensures very frequent association between voluntary

388 actions and involuntary movement. We suggest this fact lies at the heart of our finding of

389 extensible sense of agency.

390 In conclusion, we suggest that some key temporal aspects of experience of agency, namely

391 the perceptual anticipation of an action outcome, can be transferred from voluntary actions

392 to involuntary movements. Such transfer follows repeated co-occurrence of an internal

393 volitional signal, with both an involuntary body movement, and a sensory outcome.

15

394 Importantly, association with an internal volitional signal appears to be necessary to initially

395 establish key temporal features of agency with respect to an involuntary movement, but is

396 not necessary for its subsequent expression. The transfer process thus resembles the

397 development of an enduring sense of agency that emerges during skill learning, as action

398 control progresses from focussed and effortful to automatic. Interestingly, the involuntary

399 movement that becomes associated need not match the intention precisely, suggesting that

400 the metacognitive signals supporting agency acquisition are relatively non-specific. The high

401 latent associability of these signals may reflect the distributed nature of motor control.

402 Recent successes in acquisition of voluntary control using neuroprosthetics and brain-

403 machine interfaces testify to the latent associability of human sense of agency.

404 **Materials and Methods**

405 Upon arriving, participants were asked to read the information sheet and fill in the consent

406 form. In Experiment 1, participants' left index finger was attached to the distal end of the

407 robotic arm and was placed on the left control key. The right index finger was placed on the

408 enter key. The intentional binding task was explained for the participants and they were

409 familiarised with the robotic arm-induced passive movements. Two accelerometers were

410 mounted on the left and right index fingers and participants were asked to wear headphones

411 (Sennheiser, Germany). The experiment started with the baseline block (15 trials). In this

412 block, participants were instructed to look at a rotating clock but not to press any key. In

413 each trial, a tone was played and participants judged the clock hand position at the time of

414 the tone. This block was followed by the operant block, where the tone was always caused

415 by participants' keypress at a time of their own free choice, 250 ms later. Like the previous

416 block, participants were asked to judge the clock hand position at the time of the tone. In the

417 experimental group, in the first 30 trials of the operant block, voluntary keypresses of the

418 right hand were paired with the involuntary keypresses of the left hand. These *learning* trials

419 were followed by 30 *test* trials, where a command appeared on the screen and instructed

420 participants not to make any voluntary keypress with their right hand. Meanwhile, at a

16

random time, participants made a passive keypress with their left hand. As in previous trials, they made judgements about the time of the tone which followed their keypress. These test trials were interleaved with another 30 learning trials. Therefore, each operant block consisted of 60 learning trials and 30 test trials. In the control group, participants never made any voluntary action, therefore, their learning trials only consisted of passive keypress with the left hand. In both groups, the experiment finished by performing another baseline block.

Experiment 2 followed the same principles as in experiment 1 with the following differences: Participants were asked to place their left hand on the desk next to the keyboard. Robotic arm-induced movements of the left index finger were replaced with TMS-induced twitches. The TMS coil was optimally positioned in each subject to produce involuntary movement of left index finger, minimising contraction of more proximal muscles and muscles activating other joints. The headphones were replaced with loud speakers. Accelerometers were replaced with surface electrodes for electromyography (EMG). As in the previous experiment, participants made judgements about the time of the beep in three separate blocks.

Judgement error was calculated by measuring the difference between the judged clock time and the actual time. The averaged judgement error across the trials was then calculated for each block. 'Tone binding' was defined as the difference between the judgement error in the operant and the baseline condition. The negative value of tone binding represents the perceptual shift of outcome toward its action. Tone binding data from the *test* trials only were used for analysis.

Experimental design and procedure were approved by the UCL research ethics committee, and followed the principles of the Declaration of Helsinki. Transcranial magnetic stimulation followed established safety procedures (34).

References

1. Sperry RW (1940) The functional results of muscle transposition in the hind limb of the rat. *J Comp Neurol* 73(3):379–404.

2. Sperry RW (1943) Effect of 180 degree rotation of the retinal field on visuomotor coordination. *J Exp Zool* 92(3):263–279.

3. Frith CD (2014) Action, agency and responsibility. *Neuropsychologia* 55:137–142.

4. Fried I, Mukamel R, Kreiman G (2011) Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron* 69(3):548–562.

5. Desmurget M, et al. (2009) Movement intention after parietal cortex stimulation in humans. *Science* 324(5928):811–813.

6. Sherrington CS (1906) *The integrative action of the nervous system* (New York, C Scribner's sons) Available at: http://archive.org/details/integrativeacti02shergoog [Accessed October 17, 2015].

7. Prinz W (1997) Perception and Action Planning. *Eur J Cogn Psychol* 9(2):129–154.

8. Wegner DM (2003) *The Illusion of Conscious Will* (MIT Press). New Ed edition.

9. Frith CD, Blakemore, Wolpert DM (2000) Abnormalities in the awareness and control of action. *Philos Trans R Soc Lond B Biol Sci* 355(1404):1771–1788.

10. Blakemore SJ, Wolpert DM, Frith CD (2002) Abnormalities in the awareness of action. *Trends Cogn Sci* 6(6):237–242.

11. Pavlov (1927) PI (2010) Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. *Ann Neurosci* 17(3):136–141.

12. Haggard P, Clark S, Kalogeras J (2002) Voluntary action and conscious awareness. *Nat Neurosci* 5(4):382–385.

13. Moore JW, Obhi SS (2012) Intentional binding and the sense of agency: a review. *Conscious Cogn* 21(1):546–561.

14. Cravo AM, Claessens PME, Baldo MVC (2009) Voluntary action and causality in temporal binding. *Exp Brain Res* 199(1):95–99.

15. Haggard P (2008) Human volition: towards a neuroscience of will. *Nat Rev Neurosci* 9(12):934–946.

16. Synofzik M, Vosgerau G, Newen A (2008) Beyond the comparator model: a multifactorial two-step account of agency. *Conscious Cogn* 17(1):219–239.

17. Wittgenstein L (2009) *Philosophical Investigations* eds Hacker PMS, Schulte J (Wiley-Blackwell, Chichester, West Sussex, U.K. ; Malden, MA). 4th Edition edition.

18. Hart HLA, Honore T (1985) *Causation in the Law* (Oxford University Press, USA, Oxford : New York). Second edition.

19. Haggard P, Clark S (2003) Intentional action: Conscious experience and neural prediction. *Conscious Cogn* 12(4):695–707.

20. Adams RA, Shipp S, Friston KJ (2013) Predictions not commands: active inference in the motor system. *Brain Struct Funct* 218(3):611–643.

21. Moore JW, Fletcher PC (2012) Sense of agency in health and disease: a review of cue integration approaches. *Conscious Cogn* 21(1):59–68.

22. Fried I, et al. (1991) Functional organization of human supplementary motor cortex studied by electrical stimulation. *J Neurosci* 11(11):3656–3666.

23. Bays PM, Wolpert DM, Flanagan JR (2005) Perception of the consequences of self-action is temporally tuned and event driven. *Curr Biol CB* 15(12):1125–1128.

24. Skinner BF (1992) "Superstition" in the pigeon. 1948. *J Exp Psychol Gen* 121(3):273–274.

25. Koralek AC, Jin X, Long JD, Costa RM, Carmena JM (2012) Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills. *Nature* 483(7389):331–335.

26. Nicolelis MAL (2003) Brain–machine interfaces to restore motor function and probe neural circuits. *Nat Rev Neurosci* 4(5):417–422.

27. Engbert K, Wohlschläger A, Haggard P (2008) Who is causing what? The sense of agency is relational and efferent-triggered. *Cognition* 107(2):693–704.

28. Hebb DO (2002) *The Organization of Behavior: A Neuropsychological Theory* (Psychology Press, Mahwah, N.J). New Ed edition.

29. Metcalfe J, Eich TS, Castel AD (2010) Metacognition of agency across the lifespan. *Cognition* 116(2):267–282.

30. Shultz TR, Wells D (1985) Judging the intentionality of action-outcomes. *Dev Psychol* 21(1):83–89.

31. Shultz TR, Wells D, Sarda M (1980) Development of the ability to distinguish intended actions from mistakes, reflexes, and passive movements. *Br J Soc Clin Psychol* 19(4):301–310.

32. Bouisset S, Zattara M (1981) A sequence of postural movements precedes voluntary movement. *Neurosci Lett* 22(3):263–270.

33. Schieber MH (1995) Muscular production of individuated finger movements: the roles of extrinsic finger muscles. *J Neurosci Off J Soc Neurosci* 15(1 Pt 1):284–297.

34. Rossi S, Hallett M, Rossini PM, Pascual-Leone A (2009) Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clin Neurophysiol* 120(12):2008–2039.

35. Fadiga L, et al. (1999) Corticospinal excitability is specifically modulated by motor imagery: a magnetic stimulation study. *Neuropsychologia* 37(2):147–158.

36. Li S, Stevens JA, Rymer WZ (2009) Interactions between imagined movement and the initiation of voluntary movement: a TMS study. *Clin Neurophysiol Off J Int Fed Clin Neurophysiol* 120(6):1154–1160.

Figure Legends:

Figure 1.A. Timeline of an experimental trial. In the learning trials (the left green box), participants were instructed to press the enter key on a keyboard in front of them with their right index finger at a time of their own free choice. This action was paired with involuntary keypress (left control key) induced by a robotic arm pressing on the left index finger. In operant blocks, each keypress was followed by a beep (1000 Hz) after 250 ms. At the end of the trial, participants reported the perceived time of the beep. See text for full explanation. B. In the experimental group, the session started with a baseline block. The operant block then ensued. Voluntary actions of one hand were paired with involuntary movements of the other hand, followed by a tone 250 ms. After an initial learning phase of 30 trials, further learning trials were interleaved with test trials (A. right red box) on which involuntary movements were followed by tones, but no voluntary action occurred. The session ended with the execution of a further baseline block. C. A group of control participants followed the exact same design as the experimental group, but their involuntary movements were never associated with voluntary actions. In both groups, data from the test trials (red bold boxes) was used for analysis. The corresponding trial number is shown above each box. For Experiment 2, robot-induced movements were replaced with TMS-induced twitches.

Figure 2. Tone binding in test trials of Experiment 1 and Experiment 2. The dashed line indicates the perceived time of the tone in the baseline condition. Binding effects are drawn to scale, and values are in ms. Differences in baseline values across sessions have been removed for display purposes. See Figure S1 for error bars and Figure S2 for single participants' data.

Figure 3. A possible mechanism for agency transfer. A. Voluntary actions are often associated with subjective experience of volition and agency. B. In our experimental group, we rewired this association so that the participants' intention produces two movements, one voluntary and the other involuntary. C. Pairing voluntary and involuntary movements lead to

key temporal features of agency being experienced when involuntary movements were followed by outcomes (path 3). This experience, however, strongly depended on its previous association with intention which precedes the voluntary action (path 1). D. This suggests that once a voluntary signal is present, it can be mentally associated with other events, and spread to drive key temporal features of agency with respect to other movements.

Supplementary Materials

Participants. In total 72 healthy volunteers, aged 18-35 years of age, were recruited from the Institute of Cognitive Neuroscience subject data pool for separate experiments. All participants were right handed, had normal or corrected to normal vision, had no history or family history of seizure, epilepsy or any neurologic or psychiatric disorder and did not have any metallic or electronic object in the head. Participants affirmed that they had not participated in any brain stimulation experiment in the last 48 h, nor had consumed alcohol in the last 24 h. Participants were paid an institution-approved amount for participating in each session of the experiment.

Intentional binding. We used intentional binding paradigm as an implicit measure of agency. The task was based on previous studies (12), and was programmed in LabVIEW 2012 (Austin, Texas). Participants viewed a clock hand rotating on a computer screen which was located 60cm in front of the participants in a quiet room. The initial clock position was random. Clock rotation was initiated by participants pressing the return key on a keyboard. Each full rotation lasted 2560 ms. Participants were instructed to look at the centre of the clock. Depending on the trial, they made voluntary keypress by pressing the enter key with their right index finger or made involuntary keypress by pressing the left control key with their left index finger (Experiment 1). Participants chose for themselves when to make the voluntary actions. After each key press, the clock hand stopped at a random location, participants made a time judgement according to condition (see later). Each experimental session consisted of two conditions, presented in separate blocks. At the beginning of each

block, brief instructions for the relevant condition were displayed on the screen. In the baseline condition, participants were instructed to look at the clock but not to press any key. While the clock was rotating, a pure tone (1000 Hz, 100 ms duration) was played over the headphones (or a loudspeaker in the TMS experiment), 1750-4000 ms (at random) after the onset of the trial. Participants were then asked to judge the clock hand position at the time of the tone. In the operant condition, participants pressed the key at a time of their own choosing, or made an involuntary movement (depending on the trial). Each keypress (or movement) produced a tone after 250 ms. Participants had to judge the clock hand position at the time of the tone. Baseline condition was tested in two separate blocks of 15 trials, at the beginning and end of the experiment. Operant condition was tested in a single block of 90 trials between the two baseline blocks.

Haptic device. For Experiment 1, Phantom Premium 1.5 haptic device (3D Systems, South Carolina, USA) was used to induce involuntary movements in participants' left index finger. This high-precision device has 3 degrees of freedom and provides a range of motion approximating hand movement pivoting at the elbow. Distal phalanx of the participant's' left index finger was attached to the distal end of the device. Matlab 2014 (MathWorks, USA) was used to communicate with the device. The following specifications were used to induce a natural-looking passive keypress in the finger: Force direction = X:0/ Y:-1/ Z:0, force amplitude = 0.7 N, duration of downward movement = 200 ms (30 ms to taper), duration of upward movement = 200 ms (30 ms to taper). To block out the noise of the device at the time of force induction, the main body was shielded in a soundproof box and only the arm was left out through a small hole. To control for the similarity of the movements across the fingers, two accelerometers were mounted on the left and right index fingers of the participants. The kinematics of the movement were monitored by the experimenter. We measured the exact time interval between the beginning of the voluntary action and involuntary movement in the learning trials by using data from accelerometers placed on the index fingers of the left and right hand. Analysis of 10 trials selected at random showed that

it took 34 ms (± 4 ms sd) between the software command being sent to the robot and the left finger actually moving, due to the mechanical delays in the robot. Importantly, these delays are present in both learning and test trials, and in both the experimental and the control groups. The only difference between trial types is the use of a voluntary keypress to initiate the software command to the robot in the voluntary trials of the experimental group - the delay between the depression of the key and the initiation of the software command was 2 ms, and was consistent across 10 trials selected for analysis.

TMS and MEP measurement. For Experiment 2, transcranial magnetic stimulation was delivered with a Magstim 200 stimulator (Whitland, UK). The optimal location for producing twitches (Motor evoked potentials (MEPs)) in the left first dorsal interosseous (1DI) was located by systematically exploring a 1-cm grid over the hand area of the right motor cortex. The motor threshold was calculated for each subject by reducing stimulator output in 5% steps to find the lowest level at which 3 MEPs exceeding 50 μV peak amplitude were obtained from 5 successive stimulations of the relaxed 1DI. Thresholds ranged from 35% to 60% of stimulator output (mean 45%). TMS output in the experiment was set at 120% of relaxed threshold. EMG was measured from the 1DI of the left and right hand with bipolar recording from surface Ag/AgCl electrodes. These data were amplified and digitized at 2 kHz (CED 1902, Cambridge, UK).

Supplementary Results

Motor evoked potentials. Any difference in amplitude of TMS-induced twitches of the left index finger between the experimental and control group could influence tone binding. To rule out any such possibility, MEP amplitudes were compared between the two groups. No significant difference was observed between peak-to-peak amplitudes of MEPs in the test trials (t(31) = 0.54, p=0.60, d=0.19, 95% CI [-0.40, 0.69]) or in the learning trials (t(31)=-0.77, p=0.45, d=0.27, 95% CI [-0.68, 0.31]: Fig. S3) (MEP data from one subject was unavailable due to technical error). These results also exclude the possibility that participants could have produced some voluntary motor drive, however modest, in test trials – since even a minimal

24

voluntary motor command, or a 'motor image', would be expected to increase corticospinal excitability (35, 36).

Standard deviation across trials. Could the stronger intentional binding in the experimental groups reflect a confounding effect of attention? For example, when subjects perform a voluntary action, they may direct attention to that action. Any other associated event might benefit from these effects, leading to stronger perceptual learning in the experimental group than in the control group who made no voluntary actions. Such attention-enhanced perceptual learning should lead to improved time estimation for the experimental group. In fact, we found an increased bias in judgment, corresponding to stronger tone binding, in the experimental group than in the control group.

In addition, we used the standard deviation of judgement errors across trials for each participant to calculate variable error, which is inversely related to the precision of temporal estimation of the tone. Improved attention to the tone would predict lower standard deviations for the voluntary group. Standard deviation of judgement errors across trials was compared in a 2x2x2 ANOVA with the within subject factor of trial type (baseline vs. test), between subject factor of group (experimental vs. control) and the between subject factor of experiment (experiment 1 vs. experiment 2). Standard deviation across trials was higher in test trials compared to baseline trials ($F_{(1,62)}=36.61$, $p<0.01$), presumably reflecting the attentional effects of the more complex sequence of events in test conditions, particularly the co-occurrence of involuntary movement. We found no significant main effect of group ($F_{(1,62)}=0.17$, $p=0.68$), or experiment ($F_{(1,62)}=2.01$, $p=0.16$). Importantly, we found no significant interaction between trial and group ($F_{(1,62)}<0.01$, $p=0.95$), trial and experiment ($F_{(1,62)}=0.10$, $p=0.75$), or group and experiment ($F_{(1,62)}<0.01$, $p=0.98$), and no significant interaction between trial, group and experiment ($F_{(1,62)}=0.33$, $p=0.57$). Thus, we found no evidence that the experimental group had improved attention to the tone on test trials (see Fig. S4).

Supplementary figure legends

Figure S1. The perceived time of the tone (judgement error in ms) is shown in a baseline condition, where neither action nor involuntary movement occur, and in the involuntary movement test trials for the control and experimental groups. The difference between baseline and movement trials is an estimate of shift in the perceived time of the tone due to the preceding movement (grey bars). This "tone binding" effect serves as an implicit marker of sense of agency. Note inverted Y axis. Error bars show standard error of the mean. The mean (and standard error across participants) for all groups and experiments is presented in a table below the figure.

Figure S2. Tone binding in Experiment 1 (A) and Experiment 2 (B) for single participants in the experimental and control groups.

Figure S3. MEPs in the experimental and control groups, presented separately for the learning and test trials. No significant effects were observed. Error bars show standard error of the mean.

Figure S4. Standard deviation of tone judgement error across trials. Mean across participants of SD across trials is shown separately for baseline and involuntary movement test trials, and for experimental and control groups of experiment 1 (A) and experiment 2 (B). All values are in ms. Error bars show standard error across participants of the mean.

Action
Outcome
250 ms
Time
-132ms
Exp.1 Experimental group
-36ms
Exp.1 Control group
-80ms
Exp.2 Experimental group
-8ms
Exp.2 Control group

	Exp.1 Experimental	Exp.1 Control	Exp.2 Experimental	Exp.2 Control
■ Baseline trials	-31 (15)	-30 (15)	-14 (18)	-15 (16)
□ Test trials	-163 (35)	-66 (31)	-94 (29)	-23 (36)
■ Difference (tone binding)	-132 (32)	-36 (25)	-80 (25)	-8 (27)

■ Baseline trials   □ Test trials   ■ Difference (tone binding)

A.

Experiment 1

B.

Experiment 2

A.

B.