Does who I am and what I feel determine what I see (or say)? A meta-analytic systematic review exploring the influence of real and perceived bodily state on spatial perception of the external environment

Erin MacIntyre [1], **Felicity A Braithwaite** [1], **Brendan Mouatt** [1], **Dianne Wilson** [1], **Tasha R Stanton** [Corresp. 1]

[1] IIMPACT in Health, Allied Health and Human Performance, University of South Australia, Adelaide, South Australia, Australia

Corresponding Author: Tasha R Stanton
Email address: tasha.stanton@unisa.edu.au

Background. Bodily state is theorised to play a role in perceptual scaling of the environment, whereby low bodily capacity shifts visuospatial perception, with distances appearing farther and hills steeper, and the opposite seen for high bodily capacity. This may play a protective role, whereby perceptual scaling discourages engaging with the environment when capacity is low.

Methodology. Our protocol was pre-registered via Open Science Framework (https://osf.io/6zya5/) with all amendments to the protocol tracked. We performed a systematic review and meta-analysis examining the role of bodily state/capacity on spatial perception measures of the environment. Databases (Medline, PsychINFO, Scopus, Embase, and Emcare) and the grey literature was searched systematically, inclusive to 26/8/21. All studies were assessed using a customised Risk of Bias form. Standard mean differences and 95% CIs were calculated via meta-analysis using a random-effects model.

Results. 8,034 studies were identified from the systematic search. Of these, 68 experiments (3,195 participants) met eligibility and were included in the review. These were grouped into the following categories: fatigue, pain, age, embodiment, body size/body part size, glucose levels, fitness, and interoceptive accuracy. We found low level evidence (limited studies, high risk of bias) for the effect of bodily state on spatial perception. There was consistent evidence that both glucose manipulations and age influence spatial perception of distances and hills in a hypothesised direction (lower capacity associated with increased distance and hill steepness). Mixed evidence exists for the influence of external loads, embodiment, body/body-part size manipulations, pain, and interoceptive accuracy. Evidence for fitness and/or fatigue influencing spatial perception was conflicting; notably, methodological flaws with fitness and fatigue paradigms and heterogenous spatial perception measures may underlie null/conflicting results.

Conclusion. We found limited evidence for bodily state influencing spatial perception of the environment. That all studies had high risk of bias makes conclusions about reported effects reflecting actual perceptual shifts (versus merely reflecting experimental demands or error due to inadequate study design) pre-emptive. Rigorous evaluation is needed to determine whether reported effects reflect more than bias (e.g., experimental demands, inadequate blinding). Future work using reliable measures of spatial perception, comprehensive evaluation of relevant confounders, and methodologically robust (and experimentally confirmed) bodily state experimental paradigms is warranted.

1 **Does who I am and what I feel determine what I see (or**
2 **say)? A meta-analytic systematic review exploring the**
3 **influence of real and perceived bodily state on spatial**
4 **perception of the external environment**

6 Erin MacIntyre[1], Felicity A. Braithwaite[1], Brendan Mouatt[1], Dianne Wilson[1], and Tasha R.
7 Stanton[1]
8
9 [1]IIMPACT in Health, University of South Australia, Adelaide, Australia
10
11 Corresponding Author:
12 Tasha R. Stanton
13 University of South Australia,
14 IIMPACT in Health, Allied Health and Human Performance
15 GPO Box 2471, Adelaide, South Australia, 5001, Australia
16 Email: tasha.stanton@unisa.edu.au

Abstract

Background. Bodily state is theorised to play a role in perceptual scaling of the environment, whereby low bodily capacity shifts visuospatial perception, with distances appearing farther and hills steeper, and the opposite seen for high bodily capacity. This may play a protective role, whereby perceptual scaling discourages engaging with the environment when capacity is low.

Methodology. Our protocol was pre-registered via Open Science Framework (https://osf.io/6zya5/) with all amendments to the protocol tracked. We performed a systematic review and meta-analysis examining the role of bodily state/capacity on spatial perception measures of the environment. Databases (Medline, PsychINFO, Scopus, Embase, and Emcare) and the grey literature was searched systematically, inclusive to 26/8/21. All studies were assessed using a customised Risk of Bias form. Standard mean differences and 95% CIs were calculated via meta-analysis using a random-effects model.

Results. 8,034 studies were identified from the systematic search. Of these, 68 experiments (3,195 participants) met eligibility and were included in the review. These were grouped into the following categories: fatigue, pain, age, embodiment, body size/body part size, glucose levels, fitness, and interoceptive accuracy. We found low level evidence (limited studies, high risk of bias) for the effect of bodily state on spatial perception. There was consistent evidence that both glucose manipulations and age influence spatial perception of distances and hills in a hypothesised direction (lower capacity associated with increased distance and hill steepness). Mixed evidence exists for the influence of external loads, embodiment, body/body-part size manipulations, pain, and interoceptive accuracy. Evidence for fitness and/or fatigue influencing spatial perception was conflicting; notably, methodological flaws with fitness and fatigue paradigms and heterogenous spatial perception measures may underlie null/conflicting results.

40   **Conclusion.** We found limited evidence for bodily state influencing spatial perception of the

41   environment. That all studies had high risk of bias makes conclusions about reported effects

42   reflecting actual perceptual shifts (versus merely reflecting experimental demands or error due to

43   inadequate study design) pre-emptive. Rigorous evaluation is needed to determine whether

44   reported effects reflect more than bias (e.g., experimental demands, inadequate blinding). Future

45   work using reliable measures of spatial perception, comprehensive evaluation of relevant

46   confounders, and methodologically robust (and experimentally confirmed) bodily state

47   experimental paradigms is warranted.

Introduction

49   Historically, it was assumed that perception was veridical – that what one sees is an

50   accurate representation of the external word. However, contemporary evidence challenges this

51   assumption, instead suggesting that perception is an adaptive and emergent process constructed

52   from complex and dynamic interactions between one's body and the environment (Witt 2020).

53   Inference models of perception posit that past-experience and future action goals continuously

54   reshape attention and perception (Clark 2013; Clark 2018). Here, perception is a hierarchical and

55   generative process whereby top-down "predictions" aim to match (and therefore cancel out)

56   incoming sensory input. Any residual "prediction error" continues upward for further processing.

57   This flow of information is bidirectional, whereby predictions also flow downward to refine

58   incoming sensory inputs. These predictions are theorised to be optimised by the free energy

59   principle, where to maintain homeostasis, an organism must minimise the difference between its

60   generative model and perceptual inferences in its interactions with the environment, and thus

61   minimise surprise (Friston 2010). In this model, action serves to minimise sensory prediction

62   errors.

63   Embodied perception emphasises that perception is intrinsically related and grounded to

64   one's sensory-motor capacity – perception is thus a bidirectional interaction of the person as a

65   whole and their external environment (Friston 2011). Action is key to many theories of

66   perception, as movement and action are considered the interface between perception and the

67   environment. Action both drives perception (looking in a direction determines what you see) and

68   is driven by perception (turning your head towards a sound). Thus, action-specific accounts of

69   perception, which frame perception of one's environment in terms of their ability to act within it,

70   may give us insight into the mechanisms underlying perception. Based on the action-specific

71   account, the Economy of Action (EoA) hypothesis proposes that perception acts like a

72   "biological ruler", in that visual perception plays a role in scaling the external environment in

73   line with an individual's body capacity (Proffitt 2006). That is, hills look steeper when you are

74   fatigued or carrying an external load (Bhalla & Proffitt 1999) and distances appear longer if you

75   have low blood glucose (Cole & Balcetis 2013). Such perceptual shifts are thought to encourage

76   the appropriate action given one's current bodily state/capacity (e.g., explore vs. exploit). By

77   understanding the relationship between bodily state and visuospatial perception of the

78   environment, one can gain insight into ways through which perception may be scaled.

79   There is considerable criticism of the action-specific account of perception (Durgin et al.

80   2009; Firestone 2013). Several studies have failed to replicate significant effects (Durgin et al.

81   2009; Durgin et al. 2012; Hutchison & Loomis 2014), leading authors to conclude that initial

82   results may be due to alternate theories. As perception itself cannot be directly measured,

83   research relies on assessing *judgements,* and from these assessments infer the underlying

84   perception (Firestone 2013). However, as judgements are influenced by post-perceptual

85   processes (Firestone & Scholl 2017), the observed results may be due to influences other than

86   physical demands. For instance, participant's responses may be influenced by knowledge effects

87   (e.g., an alpine skier estimating hill steepness) or by experimental demands (e.g., a participant

88   correctly guessing the hypothesis of the study and altering their response accordingly) (Dean et

89   al. 2016; Durgin et al. 2012). Indeed, experiments that manipulate the participant's knowledge of

90   experimental demands (e.g., providing cover stories) have supported that social and contextual

91   factors can influence visuospatial perceptual outcomes (Keric & Sebanz 2021). This has led to

92   repeated calls for robust experimental methodology to minimise such influences (Firestone &

93   Scholl 2017; Witt 2020). However, little is known about how well current research been able to

94   integrate these calls into study methodology.

95   A recent review (Molto et al. 2020) explored the effect of action-constraints (i.e., specific

96   to research evaluating external constraints, such as tool use) on visual perception and was unable

97   to support the influence of experimental demand bias on measures of spatial perception. They

98   specifically assessed whether experimental demand bias was present through moderator analysis

99   of two commonly proposed sources of such bias – experimental design (within group effects

100   likely bigger than between group effects) and type of spatial perception used (verbal measures

101   likely to produce larger effects than visual or action-based measures). They found no evidence to

102   support that either moderator contributed to experimental demand bias. However, such analysis

103   may be pre-emptive, given that the authors did not critically appraise the overall risk of

104   experimental bias of included studies. That is, null effects may reflect high heterogeneity of

105   included studies induced by high levels of bias across many domains rather than evidence for or

106   against experimental demand bias. Therefore, here we aimed to extend this past work by

107   comprehensively exploring the effects of bodily state on spatial perception by including both

108   modifiable *and* non-modifiable bodily states, and by formally evaluating methodological risk of

109   bias. This risk of bias assessment is essential to extend pooled findings beyond simple replication

110   to more nuanced interpretation. Thus, this systematic review and meta-analysis aims to

111   summarise and critically appraise the current evidence for the effect of bodily state on measures

112   of spatial perception of the environment.

113

Methodology

115   Review methods were guided by the 2020 PRISMA guidelines (Page et al. 2021). The PRISMA

116   2020 checklist specific to our study has been completed and made available on Open Science

117   Framework (OSF) (https://osf.io/6zya5/). This review was pre-registered via uploading a locked

118   protocol to OSF on 22/01/2021 (https://osf.io/6zya5/), prior to data extraction or analysis. All

119   amendments to this protocol were tracked and uploaded to OSF.

120   *Eligibility criteria*: There was no restriction on date or language of publication. For inclusion,

121   studies were required to recruit adults (≥18 years), recruit a clinical or healthy population, and

122   evaluate the effect of bodily state on an environmental spatial perception measure. Bodily state

123   was considered to relate to physiological or morphological characteristics of participants. These

124   states could be stable (e.g., current body size), fluctuating (e.g., blood glucose levels), or

125   perceived (e.g., self-rated fitness levels). States related to affect (e.g., fear) or cognition (e.g.,

126   executive function) were excluded. Study designs could be experimental or observational, using

127   between- or within-group comparisons, or association designs; case studies were excluded.

128   Spatial perception measures of the environment (e.g., distance estimations or hill steepness

129   estimations) were included, but measures of peri-personal space (including line bisection tasks to

130   delineate boundaries between peri-personal and extra-personal space) and perception of object

131   size were excluded. See Supplementary 1 for full details.

132   *Literature search:* A systematic search of Medline (OVID), PsychINFO (OVID), Scopus

133   (Elsevier), Embase (OVID), and Emcare (OVID) was undertaken from inception to 26/8/21.

134   Keywords and subject headings related to both spatial perception and bodily states were used

135   (Supplementary 2 provides the full search strategy).

136   Consistent with best-practice guidelines (Higgins et al. 2021), a grey literature search was

137   conducted with non-peer reviewed studies considered eligible for inclusion in the review.

138   Inclusion of grey literature was purposeful to help mitigate against publication bias and inflated

139   effect sizes. Past work has shown that excluding grey literature can result in exaggerated effect

140   sizes (McAuley et al. 2000). ProQuest Theses and Dissertations was searched using a similar

141   search string, and Google Scholar was searched using ten strings of keywords, with the first 20

142   results of each search screened. Abstract titles from relevant conferences (Vision Sciences

143   Society; European Conference on Visual Perception) were screened using keyword searches.

144   Citations of all included studies were also screened for potentially relevant studies by the

145   primary author.

146   ***Study Selection:*** Records identified by the search were exported to EndNote (Clarivate

147   Analytics, Philadelphia) and duplicates were removed. The remaining records were uploaded to

148   Covidence (Veritas Health Innovation Ltd., Australia) for screening. All title and abstracts were

149   screened by two independent reviewers (EM, BM, FB, DW). Full texts of potentially relevant

150   studies were then assessed for inclusion, again by two independent reviewers. Any

151   disagreements were resolved through consensus discussion, and if needed, a third reviewer (TS)

152   was consulted.

153   ***Risk of Bias Assessment:*** Risk of bias was assessed using a customised, piloted tool based on the

154   Appraisal tool for Cross-Sectional Studies (AXIS) (Downes et al. 2016), with modifications to

155   allow appraisal of a range of study designs. The tool consisted of 15 items evaluating participant

156   selection, study methodology, blinding, and reporting (See Supplementary file 3 for the full tool

157   and scoring instructions). Two independent reviewers (EM, and BM, FB, or DW) appraised each

158   study and discrepancies were resolved through discussion, and if required, via consultation with

159   a third reviewer (TS).

160   ***Data extraction:*** Two independent reviewers (EM and BM, FB, or DW) extracted the following

161   data using a piloted, customised form: source details (authors, year), study design, study sample

162   (bodily state examined, assessment method), participant information (sample size, age, and

163   gender), presence of a control group/condition (if applicable), type of experimental manipulation

164   (if applicable), the type of spatial perception measures used, manipulation check for bodily state

165   (if applicable) and spatial perception measure results (measures of central tendency and

166   dispersion or association). Discrepancies were resolved through discussion then, if required,

167   through consultation with a third reviewer (TS). When outcome data were missing or reported in

168   a form unamenable to quantitative meta-synthesis, authors were contacted a maximum of three

169   times, after which the data were considered irretrievable (and included only in descriptive

170   synthesis). In situations where data were presented only graphically and were not available from

171   the authors, data were extracted by two independent reviewers (EM, and BM, FB, or DW) using

172   WebPlotDigitizer (Rohatgi 2020), which has high inter-coder reliability and validity (Drevon et

173   al. 2017). If >10% discrepancy between extractors occurred, data were extracted again by a

174   single reviewer with the average of the two closest extractions used.

175   ***Data Analysis and Synthesis*:** Studies/experiments were grouped into the following bodily state

176   categories by the primary author: fatigue (sleep deprivation, pre/post exercise manipulations),

177   pain, age, body/body part size manipulations, embodiment (ownership of a virtual body or a

178   wheelchair), fitness, glucose manipulations, external loads (backpacks, ankle weights), and

179   interoceptive accuracy. We considered such subgrouping necessary, as previous work (Molto et

180   al. 2020) found that different constraint manipulations (tool-use, external loads, and effort) had

181   different effects on measures of spatial perception (Hedge's g ranged from 0.14-0.4), which may

182   represent different mechanisms through which spatial perception is scaled. Additionally, bodily

183   state comparisons/manipulations had differing risks of bias associated with the

184   comparison/manipulation. Given that studies with high risk of bias often have inflated effect

185   estimates, subgrouping was necessary to avoid overall inflation of pooled estimates driven by

186   certain bodily state comparisons/manipulations (Schäfer & Schwarz 2019).

187   Our registered protocol proposed broad groupings of bodily state: i) physiological, e.g.,

188   glucose manipulation; ii) morphological, e.g., actual body weight/size; and iii) perceived, e.g.,

189   altered perception of the body without actual change, such as via virtual reality. While our

190   chosen groupings are relatively consistent with the original protocol, the large number and

191   variety of bodily state manipulations, with varying risk of bias of the manipulation, resulted in

192   the need for further groupings. Given this deviation from original protocol, to mitigate

193   confirmation bias, an assessor blinded to the study findings (TS) reviewed the groupings and

194   made recommendations regarding pooling – i.e., which outcome measures from included

195   studies/experiments were sufficiently similar for quantitative synthesis.

196   *Meta-analysis method:* Pooling was undertaken when data were available for at least two

197   studies/experiments that were sufficiently similar (population and/or manipulation of bodily

198   state, type of spatial perception measure used). Separate forest plots were completed for each

199   spatial perception measure in each bodily state grouping, with subgroups based on the magnitude

200   of spatial perception measures (e.g., 4m, 8m 12m for distance estimations). Studies/experiments

201   using similar distances [+/- 10%] were placed in the same subgroup. An overall pooled effect for

202   all subgroups was considered when: i) pooling did not contradict the study authors' hypotheses

203   (e.g., if an effect of short vs far distances was expected, overall pooling was not undertaken to

204   avoid missing potential nuanced effects); ii) there was sufficient variability of studies/

205   experiments within each subgroup, with adequate sample sizes. If a study/experiment was

206   included in numerous subgroups (e.g., multiple distance estimations), the required commensurate

207   reduction in sample size based on the number of times it appeared in the forest plot was

208   undertaken, which can make data from small experiments meaningless due to reduced sample

209   size.

210   Where sufficient data were provided, standard mean difference (Hedge's d) was

211   calculated to provide a measure of effect size. For between group designs the following formula

212   was used (Borenstein et al. 2011).

213
$$d = \frac{\overline{X}_1 - \overline{X}_2}{S_{within}}$$

214   Where S_{within} is calculated by:

215

216
$$S_{within} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}}$$

217

218   For within group designs Hedges d was calculated using the same formula, but using an

219   appropriate calculation for S_{within} to account for non-independence of the groups (Borenstein et

220   al. 2011):

221
$$S_{within} = \frac{SD_{diff}}{\sqrt{2(1 - r)}}$$

222

223   The variable r is the correlation between the within subject data points. Where r was not

224   reported or could not be calculated with available data, we imputed a correlation value from a

225   similar study (with similar comparisons); in these cases, sensitivity analyses were performed

226   using correlations ±0.1, consistent with best practice (Higgins et al. 2021).

227   The majority of included studies (N=44, 68%) had a sample size of less than 50 and the

228   largest sample was 269. Therefore, because d is biased in experiments with small sample sizes

229   (Hedges 1981), for consistency and ease of comparison, we computed Hedge's g for all studies

230   by multiplying Hedge's d by correction factor J (Borenstein et al. 2011).

231
$$J = 1 - \frac{3}{4df - 1}$$

232   Standard error was calculated with the following formulas for between and within group

233   designs respectively (Borenstein et al. 2011).

234
$$SE = \sqrt{\frac{n_1 + n_2}{n_1 \times n_2} + \frac{d^2}{2(n_1 + n_2)}}$$

235

236
$$SE = \sqrt{\left(\frac{1}{n} + \frac{d^2}{2n}\right) \times 2(1 - r)}$$

237   For experiments that evaluated the association between two continuous variables,

238   Fisher's z and its variance were calculated from the correlation using the following formulas

239   (Borenstein et al. 2011):

240
$$z = 0.5 \times \ln\left(\frac{1 + r}{1 - r}\right)$$

241   and

242
$$V_z = \frac{1}{n - 3}$$

243   Where experiments evaluating the same bodily state used differing outcomes (e.g.,

244   correlations versus means), r was converted to d and the variance of d was calculated using the

245   following formulas (Borenstein et al. 2011) to allow pooling, before being converted to g with

246   the correction factor J.

247
$$d = \frac{2r}{\sqrt{1 - r^2}}$$

248
$$V_d = \frac{4V_r}{(1 - r^2)^3}$$

249   Where V_r was computed with the following formula:

250 $$V_r = \frac{(1 - r^2)^2}{n - 1}$$

251   When linear mixed models (LMM) were used and the original dataset was unavailable,

252 the results were only presented narratively, consistent with recent recommendations (Roth et al.

253 2018). Specifically, given that effect sizes derived from LMM are dependent on the other factors

254 included in the model, they should only be pooled when other studies also include the same

255 factors within their analysis.

256   Calculated effect size and SE were imported to R (Team 2021) using the Metafor and Meta

257 packages (Balduzzi et al. 2019; Viechtbauer 2010). A random-effect, generic inverse variance

258 meta-analysis was undertaken (Higgins et al. 2021). Where meta-analysis was not possible, the

259 calculated effect size and 95% CIs are reported separately for each study. To assess potential for

260 publication bias, contour enhanced funnel plots (Egger et al. 1997) were created for each forest

261 plot, using the funnel.meta function, with contours set at $p < 0.1$, $p < 0.05$, and $p < 0.01$.

262

263 **Results**

264 ***Search Results:*** A total of 8034 records were identified using the database search strategy. Of

265 those, 2536 were duplicates, resulting in 5498 unique records screened for inclusion, with full

266 text review of 94 records (containing 179 individual experiments). From this, 62 studies (132

267 experiments) were excluded (See Figure 1), resulting in 32 studies (47 experiments) included.

268 An additional 15 studies (21 experiments) were included through the grey literature search,

269 resulting in a total of 47 studies (68 experiments). A list of all studies that underwent full text

270 review and the reasons for their exclusion has been included in Supplementary File 4. Of these,

271 14 experiments had not been peer-reviewed (conference abstracts or dissertations). Forty-four

272 experiments had sufficient data to allow potential inclusion in the quantitative analysis (n=25 via

273 graphical extraction). Additional data were obtained for 22 experiments via contacting the

274 authors. In two experiments, data could not be obtained, thus effect sizes could not be calculated.

275 Overall, 53 experiments (from a total of 37 studies) were included in the quantitative analysis.

276 Table 1 provides methodological details of these included experiments.

277

278 ***Characteristics of included experiments***

279 Study design: The majority of experiments (n=42/68) used between group designs, either

280 observational (e.g., young vs. old age groups), or via experimental manipulation (e.g., glucose

281 vs. placebo glucose groups). Thirteen experiments used repeated measure within group designs,

282 and six experiments used mixed designs (within x within repeated measures; or between x within

283 designs). Eight experiments used a cross-sectional design to explore associations between bodily

284 state and spatial perception.

285

286 Spatial perception outcomes: All studies assessed either distance or hill/staircase steepness

287 estimations. Estimated distances were to targets, or of aperture or gap size. Estimated distances

288 ranged from 0.5m- 92.5m, and estimated hill slopes from 2deg-39deg. One study (Tenhundfeld

289 & Witt 2015) evaluated 'distance on hill' (difference between estimations for a distance on flat

290 ground and the same distance on a hill). The methods used to evaluate spatial perception were

291 variable but were broadly categorized into verbal, visual (visual matching in the environment or

292 on a diagram), haptic (participants used their hand to match the slope of the hill), or action-based

293 (blind walking, beanbag toss).

294 ***Overall Risk of Bias:*** All experiments had high risk of bias (Table 2). No experiments used

295 random sampling and only 4% (N=3/68) performed *a priori* sample size calculations (Scandola

296 et al. 2019; Shaffer et al. 2019). Four experiments (Dean et al. 2016; Durgin et al. 2012; Shaffer

297 et al. 2013; Taylor-Covill & Eves 2013) had low risk of bias for the spatial perception measure

298 (i.e., reliability established by previous research). Only 22% (N=15/68) of experiments

299 adequately controlled for confounding variables (such as controlling for time of day or previous

300 night's sleep in fatigue paradigms). Blinding was poorly applied: 21% (N=14/68) had a low risk

301 of bias for blinding participants (participants were blinded and this was assessed with post-

302 experiment questioning), 3% (N=2/68) blinded assessors, and no experiments reported a blinded

303 analysis. None of the included experiments were pre-registered. Figure 2 illustrates the

304 percentage of experiments that scored high, low, unclear, or N/A for each risk domains.

305

306 ***Outcomes***: ***Effect of bodily state on spatial perception***

307 Effect of fatigue on spatial perception (N=8): All experiments aimed to induce fatigue through

308 exercise, comparing measures of spatial perception either pre-and post-exercise, or between

309 exercise and non-exercise groups. Two experiments used a fatigue manipulation via sleep

310 deprivation in addition to an exercise manipulation.

311 *Distance estimation (N=5):* All experiments evaluated verbal distance estimation data. Pooling

312 of two experiments (Baati et al. 2020; Baati et al. 2015) that compared distance perception pre-

313 vs post-exercise using a sprint cycling task (10x 6sec maximal cycling) found no effect of

314 exercise at any of the three distances measured, for either the normal (Figure 3a) or the sleep

315 deprived conditions (Figure 3b). Within the same studies, there was an effect of fatigue due to

316 sleep deprivation only at the 35m distance (Figure 3c), where participants perceived the distance

317   as *closer* when undergoing sleep deprivation compared to a normal night sleep. A final study

318   (Hunt et al. 2017) evaluated the effect of fatigue (stepper machine 90 sec) versus non-fatigue,

319   finding that the fatigued group estimated a single target distance of 92.5m as being farther than

320   the non-fatigued group (Hedge's g: 0.56, 95%CI -0.95 to -0.16).

321   Two experiments used comparable distance measures, but in athletic populations (Asaf et

322   al. 2015; Jarraya et al. 2013). Asaf et al. (2015) showed no difference (Hedge's g range: -0.13 to

323   0.09) in distance estimations based on their fatigue manipulation (treadmill running 2 min),

324   whereas Jarraya et al. (2013) found very large effect sizes following cycling for 10 mins

325   (Hedge's g range: 8.32 to 10.08), indicating participants estimated distances as *closer* in the post-

326   exercise (fatigued) condition compared to the pre-exercise condition. Pooling of the studies'

327   similar distances (12m and 11m) was non-significant, with high heterogeneity ($I^2 = 96\%$). Large

328   differences in effect sizes at a similar distance (g= -0.14 vs 9.92), suggest that between study

329   differences are unlikely driven by differences in the spatial perception task; but rather, may lie in

330   the fatigue manipulation. Sensitivity analyses evaluating the influence of varying imputed

331   correlation coefficients showed similar results (Supplementary File 5, Tables 1-4).

332   *Hill steepness estimation (N=3):* Three experiments (Bhalla & Proffitt 1999; Proffitt et al. 1995;

333   Taylor-Covill & Eves 2013) evaluated the effect of fatigue (all running to exhaustion/maximal

334   fitness tests) on steepness estimation of hills (n=2) or stairs (n=1) and were able to be pooled.

335   For verbal measures of steepness (Figure 3d), participants estimated hills as steeper after

336   fatiguing exercise for both shallow hills (5deg) and steep slopes (15-31deg). For haptic measures

337   of steepness (Figure 3e), participants also estimated shallow hills as steeper after fatiguing

338   exercise, but not steep hills. Sensitivity analyses, as described above, showed similar results (See

339   Supplementary File 5, Tables 5 and 6).

340   *Interim discussion:* There was inconsistent and conflicting evidence for an effect of fatigue on

341   spatial perception of distance, although consistent evidence was seen for hill steepness, with

342   fatigue resulting in reports of increased hill steepness.

343   For distance spatial perception, pooling found no effect of fatigue, as induced by exercise

344   (Figure 3a), exercise and sleep deprivation (Figure 3b) or sleep deprivation alone (Figure 3c).

345   However, our risk of bias appraisal identified potential issues with fatigue paradigms (e.g.,

346   insufficient fatigue for perceptual scaling to occur) and/or distance estimation tasks (e.g.,

347   distances not sufficiently challenging for the population). For example, in both pooled studies

348   (Baati et al. 2020; Baati et al. 2015) that found null results, a relatively easy sprint cycling task (6

349   seconds of sprinting x 10), particularly for a young population (mean=23 years), was used to

350   induce fatigue. In contrast, Hunt et al. (2017) recruited an older population (mean=54 years),

351   used a challenging exercise task for that population (stepper task), and used longer distances

352   (92.5m), and did show findings consistent with EoA hypothesis: fatigued participants estimated

353   distance as further than non-fatigued participants. Exercise task difficulty in context of the

354   population may also have relevance to findings (Jarraya et al. 2013) that were *contrary* to what

355   EoA would predict (fatigued athletes estimated distances as being *closer* than when they were

356   not fatigued). When a highly trained athletic population undertakes a task that is not sufficiently

357   challenging to induce fatigue, it is possible that this increases activation/arousal and perceived

358   energy/vigour (Thayer 1978), inducing an opposite perceptual shift. Given that the studies did

359   not assess whether their fatigue manipulations were successful, it is difficult to know if they were

360   adequate to induce perceptual scaling consistent with the EoA. However, these inconsistent

361   findings may also provide evidence *against* EoA. Indeed, given high risk of bias for participant

362   blinding and given the populations recruited (e.g., often athletes vs non-athletes), social effects

363   on compliance with experimental demands and knowledge effects cannot be ruled out. In

364   particular, athletes (such as the soccer players in both Baati et al. [2015; 2020] experiments),

365   may have substantial knowledge/experience in estimating distances. Thus, we cannot rule out the

366   effect of such knowledge on the results.

367   The "difficulty" of the spatial perception task may also support more consistent findings

368   seen here for hill steepness estimation. Hills are more difficult to walk than flat surfaces (hills

369   requiring more capacity), thus hill estimation may require less 'fatigue' to induce a perceptual

370   shift than flat distance measures. Indeed, pooling of verbal estimations of hill steepness found

371   that people who were fatigued typically estimated hills as being steeper than control

372   groups/conditions (consistent with EoA), although findings were less clear when using haptic

373   measures. Further work is needed to clarify such findings, particularly given that many pooled

374   studies were small, conducted by the same research groups, and the actual fatigue level induced

375   was rarely evaluated.

376

377   Effect of pain on spatial perception (N=3):

378   *Distance estimation (N=3):* Two studies compared walking distance estimation in people with

379   and without persistent pain (Tabor et al. 2016; Witt et al. 2009) were pooled. There was no

380   significant effect of pain on distance estimation (Figure 4). High heterogeneity was seen (e.g.,

381   73%, 83%), where Witt et al. (2009) found that people with lower limb pain verbally estimated

382   distances as farther than pain-free controls, and Tabor et al. (2016) found no differences between

383   pain and pain-free groups. The third study (Alaiti et al. 2019) evaluated the influence of shoulder

384   pain on ratio judgements during a pointing task, where participants estimated the distance to

385   various points on a board (see Table 1 for details). There was no significant effect of shoulder

386 pain on distance estimation (g = -0.33, 95% CI -0.68 to 0.02), but a trend towards those with

387 shoulder pain estimating distance as farther than those without shoulder pain was present.

388 *Interim discussion:* There was limited and inconsistent evidence for an effect of pain on distance

389 perception. However, the risk of bias assessment identified the potential importance of pain by

390 task interactions i.e., that the location of body pain might need to be relevant to the actual

391 performance of the spatial perception task. Participants in Witt et al. (2009) had lower limb pain

392 and showed evidence of perceptual scaling for walking distance estimations. In contrast, Tabor et

393 al. (2016) recruited people with a variety of chronic pain conditions in various body locations,

394 including those whose pain might not interfere with the person's capacity to traverse the

395 environmental distances presented to them, and had null findings. That the spatial perception

396 task might require specificity to the site of pain (i.e., site of reduced bodily capacity), is

397 supported by results found by Aliati et al. (2019) of a trend towards pain-induced effects on a

398 distance perception task that was specific to the area of pain.

399   Interestingly, a secondary analysis within the included study by Tabor et al. (2016)

400 showed no distance estimation differences between those that anticipated experiencing pain

401 during walking versus those that did not; it is likely that this post hoc analysis was

402 underpowered. Alternatively, pain may be a proxy for another bodily state – e.g., pain-related

403 functional limitation that results in reduced cardiovascular fitness – that interacts with pain or

404 solely drives changes to spatial perception. Our risk of bias assessment found that none of the

405 studies in this grouping controlled for any confounding variables. Therefore, future studies

406 evaluating numerous putative contributors to spatial perception are warranted.

407 Effect of age on spatial perception (N=10):

408   *Distance estimations (N=8):* Six experiments evaluated verbal estimations of distance and were

409   sufficiently similar to be pooled (Bian & Andersen 2013; Costello et al. 2015; Sugovic & Witt

410   2013). This meta-analysis revealed that older participants estimated distances as farther than

411   younger participants, at all five distances examined (Figure 5a). There was also a negative effect

412   of age on distance perception at three out of five distances in the single experiment that used an

413   action-based measure (Bian & Andersen 2013); see Figure 5b. An equidistant cone task was

414   nonsignificant at all distances (Norman et al. 2020).

415   *Hill steepness estimation (N=2):* Two experiments (Bhalla & Proffitt 1999; Dean et al. 2016)

416   examined the effect of age on measures of hill steepness but were unable to be pooled. They both

417   found a positive relationship when estimating the steepness of moderate hills (e.g., 9 or 10

418   degrees), indicating that the elderly group estimated the hills as *less steep* than the young group.

419   In an LMM, Dean et al. (2016) reported a beta coefficient for age of -0.2deg/year, indicating that

420   hill steepness estimates decreased by 0.2deg every one year of increased age. However, while

421   Bhalla & Proffitt (1999) also found that older participants estimated the hills as *less steep* than

422   younger participants for shallow hills (see Figure 5c), the opposite was seen for steeper hills (25

423   or 29 degrees) where older participants estimated hills as *more steep* than younger participants.

424   *Interim discussion:* While there is consistent evidence that age shifts distance perception in line

425   with the EoA (older participants perceive distances as farther than younger participants), the

426   evidence for hill steepness perception is limited and less consistent. Similar to findings in fatigue

427   studies, a potential task-related influence may be present whereby shallow hills may not be

428   sufficiently "difficult" to induce a challenge to homeostatic state, and thus, show an effect of age,

429   whereas steep hills do show an effect of age. Findings from pooled distance estimations are

430   supportive of an influence of 'difficulty:' older participants over-estimate distance to a greater

431 extent than younger participants [12 m; g= -1.48 (95% CI -2.04 to -0.91)] and this age-induced

432 difference is smaller for shorter distances [4m; g= -0.75 (95% CI -1.19 to -.30)]. However, this

433 does not fully explain findings that older individuals perceive shallow hills as less steep than

434 younger individuals.

435   Potential limitations exist for pooled distance estimations with four of the six experiments

436 from the same paper (although four different participant groups). While this raises the possibility

437 of systematic error biasing the pooled result, the observed effects in the four experiments were

438 smaller than those found in the others. Thus, if a systematic bias was present, it would be in the

439 direction of a null effect, which provides increased confidence that findings of an effect of age

440 on distance perception are robust. Our risk of bias assessment indicates that potential

441 confounders were poorly controlled in this grouping (all but (Dean et al. 2016) [90%] scoring

442 high or unsure in this risk of bias domain). Therefore, the observed differences between groups

443 may be due to knowledge effects – where older participants have greater experience in hill and

444 distance estimations than younger participants. Conversely, the between group study designs

445 employed in this grouping may have reduced response bias, as participants would be less likely

446 to guess the study's hypothesis/anticipated direction of effect.

447

448 Effect of body size/part manipulation on spatial perception (N=7):

449 *Distance/aperture width estimations (N=4):* Two experiments evaluated the effect of increased

450 hand size (via padded gloves) on a visual matching aperture width measure (Collier 2018) were

451 pooled, finding no effect(Figure 6a). Contrasting results were seen when foot size was

452 manipulated in virtual reality (Jun et al. 2015): those in the large foot group estimated gaps as

453 smaller than those in the small foot group. Similarly, Van der Hoort et al. (2011) used augmented

454   reality to perform a "body swap" and found that taking on a small body resulted in over-

455   estimation of distance, compared to a normal size body, using both verbal and blind-walking

456   measures. Further, taking on a large body led to underestimation of distance (Figure 6b).

457   Sensitivity analyses evaluating the influence of varying imputed correlation coefficients showed

458   similar results (See Supplementary File 5, Table 7).

459   *Body size manipulations, Hill steepness estimations* (N=2): Two experiments from one study

460   (Bridgeman & Cooke 2015) compared hill steepness estimation between conditions of standing

461   on the ground versus standing on a 37cm box. Pooled estimates found a positive effect on hill

462   steepness where participants standing on a box estimated that the hill was less steep than

463   participants standing on the ground, at 2m and 4m, but not at 8m or 16m (see Figure 6c).

464   *Interim discussion:* While evidence for an effect of body size manipulation on spatial perception

465   was inconsistent, this finding likely reflects high heterogeneity in body manipulation methods,

466   including potential use of inadequate manipulation paradigms. Three of the seven experiments

467   found an effect that supports the EoA at every distance evaluated (and for both verbal and action-

468   based measures), whereby increasing body size decreased perceived distance, and decreasing

469   body size increased perceived distance. Of these, two experiments assessed body size via

470   embodiment of VR avatars and did determine that their VR experimental manipulation was

471   successful (participants embodied the large/small bodies). The remaining four experiments had

472   null or mixed results, although lack of manipulation checks limits the ability to interpret their

473   findings. For example, Collier (2018) found that padded gloves (vs. no gloves) did not change

474   spatial perception. Given the lack of a manipulation check, it is unclear whether use of padded

475   gloves solely influenced perceived hand size, whether it resulted in a feeling of encumbrance, or

476   whether it did not alter bodily perception at all. All studies in this grouping scored high or unsure

477   on participant blinding in the risk of bias assessment. Given that most studies in this grouping

478   used within-group designs, lack of blinding may lead to response bias (i.e., participants guessing

479   the hypothesised direction of the effect after being exposed to both manipulation conditions),

480   reducing confidence in these studies' results. Finally, limitations in the experiments manipulating

481   eye height (via participants standing on a box) make it difficult to determine whether the results

482   provide evidence against EoA or inconsistent results are due to task constraints. The lack of

483   consistent effect on estimated hill steepness for *all* distances was interpreted as lack of support

484   for EoA (all points on a hill being equally difficult to climb). However, that repeated verbal

485   estimates of the *same, unchanging physical* hill were used, suggests possible response bias

486   contribution.

487

488   Effect of embodiment on spatial perception (N=3)

489   *Distance estimations (N=3):* Two experiments (Phillips et al. 2010; Ries et al. 2008) compared

490   distance estimation (via blind-walking) of people embodying an avatar (vs. no avatar) in virtual

491   reality. Pooling found that the avatar group estimated distances as farther than the no-avatar

492   group (Figure 7a). A third study (Scandola et al. 2019) investigated the influence of wheelchair

493   embodiment in people with spinal cord injury (SCI). There was a negative effect of embodiment

494   on distance perception (via distance on hill): people estimated distances as farther for steep hills

495   when in their own (embodied) wheelchair versus a generic (non-embodied) wheelchair; there

496   was no effect for distances on shallow hills (Figure 7b).

497   *Hill steepness estimations (N=1):*

498   Scandola et al. (2019) also evaluated the effect on hill steepness perception of people with SCI

499   and found no difference between using their own (embodied) wheelchair versus a generic

500   (unembodied) wheelchair for visual matching estimation of steepness at any hill angle (Figure

501   7c).

502   *Interim discussion:* Findings were consistent that embodiment (either of an avatar or of a

503   wheelchair) results in perceptual scaling for distance estimates (makes distances seem farther). A

504   strength of these studies was that embodiment paradigms were experimentally evaluated,

505   allowing confidence that the experimental manipulation or condition evaluated the construct it

506   intended to. Only one study (Scandola et al. 2019) explored the influence of embodiment on hill

507   steepness perception, therefore replication is needed. Two studies in this grouping (Phillips et al.

508   2010; Ries et al. 2008), were conference abstracts and scored poorly on the risk of bias

509   assessment (both scored high or unsure in all categories except for outcome reporting and

510   missing data), thus reducing confidence in the results.

511

512   Effect of glucose manipulations on spatial perception (N=10):

513   *Distance estimation (N=5):* These experiments were unable to be pooled due to methodological

514   differences, including type of distance estimation measure (action-based, via blind-walking and

515   bean-bag toss, and visual matching tasks). Four experiments found that participants in the

516   glucose group estimated distances as closer than did those in the control group (placebo glucose;

517   Figure 8a) (Cole & Balcetis 2013; Zadra 2013; Zadra et al. 2016). One experiment had a non-

518   significant result (Cole & Balcetis 2013), but the results were consistent with the overall

519   direction of effect seen in the other studies.

520   *Hill steepness estimations (N=5):* All five studies evaluated verbal hill steepness estimation

521   (Durgin et al. 2012; Schnall et al. 2010; Shaffer et al. 2013; Zadra 2013) and were able to be

522   pooled (Figure 6a). Ingesting glucose resulted in less steep verbal estimation of hill angle

523   (compared to placebo) for shallow hills, but not steeper hills. There was high heterogeneity in

524   pooled analyses of steep hills (I2 = 81%) with one study (Schnall et al. 2010) finding a large

525   effect in favour of glucose reducing perceived hill steepness, and one with null results (Shaffer et

526   al. 2013).

527   *Interim discussion:* There was consistent evidence that participants who ingested glucose

528   estimated distances as shorter and hills as less steep than in control groups/conditions – in line

529   with the EoA. While a non-significant pooled effect was found for the steep hill subgroup, the

530   study that found a small non-significant effect (Shaffer et al. 2013) was assessed to have a *higher*

531   risk of bias (did not control for confounders, and missing data >15%) than the study that did find

532   a significant effect of glucose (Schnall et al. 2010), thus more weight is placed on the latter's

533   findings. Furthermore, in all but one study (Cole & Balcetis 2013), participants ingested

534   food/drink either containing glucose or a sugar free sweetener (placebo) allowing participant

535   blinding. In addition, this work contained the only two experiments (Cole & Balcetis 2013;

536   Zadra 2013) where assessors were blinded. Adequate blinding and the use of between group

537   study designs means that competing theories (experimental demands and response bias) are less

538   likely to have influenced the results, increasing confidence that the findings do reflect an effect

539   of glucose on measures of spatial perception.

540

541   Fitness and measures of spatial perception (N=11): Experiments used either objective proxies of

542   fitness (e.g., BMI or waist-to-hip ratio), or perceived fitness (e.g., subjective ratings of fitness or

543   body size). Given the diversity of measurements, we standardised the direction of effects as

544   follows: a negative effect would indicate that *increased* actual/perceived fitness resulted in

545   *decreased* measures of spatial perception (distances shorter, hills less steep).

546   *Distance estimation (N=4):* Methodological differences prevented pooling. All effects sizes were

547   negative, but only two significant (Figure 9a). One study (Sugovic et al. 2016) evaluated the

548   effect of BMI on verbal distance estimation finding significant effects for estimates at 10m, but

549   not at any other, longer distances, as well as no effect of perceived body size (picture selection

550   task) at any distance. Taylor et al. (2011) found that traceurs/parkour athletes (i.e., higher

551   fitness/sport-specific training) estimated vertical wall height as significantly lower than untrained

552   controls for a 1.94m wall but not for higher walls (2.29m, 3.45m). Effect sizes were unable to be

553   calculated for two other studies, but one found a non-significant relationship between waist-to-

554   hip ratio and distance perception (Cole et al. 2013), and one found no relationship between self-

555   rated fitness and a distance on hill measure (Shaffer et al. 2019).

556   *Hill steepness estimations (N=7):* Three experiments (all from the same paper) that evaluated the

557   effect of self-rated fitness on verbal (Figure 9b) and haptic (Figure 9c) measures of hill steepness

558   were able to be pooled (Krpan & Schnall 2014) and produced non-significant results. The verbal

559   measure trended towards a positive effect, indicating that people who rated their own fitness as

560   higher, also perceived the hill as steeper; while the haptic measures trended in the opposite

561   direction, towards a negative effect.

562   The remaining studies used objectively measured weight/fitness and found effects on

563   perceived hill steepness (Figure 9d). Sugovic et al. (2014) found that hills were estimated as

564   steeper as BMI increased. Similarly, Bhalla et al. (1999) combined BMI and a submaximal

565   fitness test to create a composite "fitness measure", and found that as fitness increased, estimated

566   hill steepness (verbal) decreased. Haptic responses had no relationship with fitness

567   measurements. In another experiment (Taylor-Covill & Eves 2016) experimenters used weight

568   coding to group participants into "clearly overweight" and "healthy" groups, and also found that

569   overweight status increased perceived hill steepness estimates. Likewise, Dean et al. (2016)

570   found that people with higher BMI gave higher slant estimations (LMM beta = 0.5). A final

571   experiment (Taylor-Covill & Eves 2016) also found an association (via LMM) between body

572   composition (measured via dual-energy x-ray absorptiometry [DEXA] at baseline and follow-up;

573   average of 407 days between assessments) and perceived steepness of staircases. At both

574   timepoints BMI and fat mass were positively associated with verbal and visual measures

575   (increases in BMI/fat mass associated with increased estimated steepness), but the model was

576   non-significant for haptic measures of steepness. Longitudinally, when BMI and body fat

577   decreased, so too did verbal estimations of steepness, but not haptic.

578   *Interim discussion:* There was no evidence for a relationship between *self-rated* fitness and

579   spatial perception, but there was consistent evidence for associations with objective fitness

580   measures.

581   Limitations in the measurement of self-rated fitness likely contribute to the lack of effect.

582   For example, Krpan et al. (2014) assessed perceived fitness (self-rated) using a 5-point scale, and

583   had poor distribution of responses [e.g., in experiment 1, 82.7% of participants rated their

584   condition as either good (4) or excellent (5)], and thus, likely has poor sensitivity to differences

585   in perceived fitness. The perceived fitness outcome measures used by Sugovic et al. (2016) were

586   similar, with poorly defined constructs (e.g., body type category of Large/Muscular) and limited

587   response options. Unclear validity of the fitness measure cannot be ruled out as a contributor to

588   the lack of relationship between fitness and distance measures.

589   While more consistent evidence in line with EoA was present for objective measures of

590   fitness, such as BMI, waist-to-hip ratio, and fat mass influencing spatial perception, methods of

591   assessing objective fitness were also problematic. Authors reported measuring "physical

592   abilities" via BMI (Sugovic et al. 2016) and "physiological fitness" via waist-to-hip ratio (Cole

593   et al. 2013). While these measures hold correlations with measures of fitness (Wier et al. 2006),

594   they have important limitations that may be relevant to EoA. For example, BMI does not account

595   for differences between muscle and fat. Thus, a heavily muscled, but fit individual (e.g., Lebron

596   James, BMI 27.5; overweight category), would have a higher BMI then a sedentary individual,

597   with low muscle mass and high body fat, but we would expect very different results for spatial

598   perception based on the EoA hypothesis (Nevill et al. 2006). One study evaluated "body

599   composition" (Taylor-Covill & Eves 2016) using DEXA scanning, which does allow for the

600   delineation between muscle and fat. Indeed, this study found that fat mass, rather than fat free

601   mass, was significantly associated with measures of spatial perception both cross-sectionally and

602   longitudinally. Additionally, consistent with other groupings, these studies scored high risk of

603   bias for the blinding domain. Thus, it is difficult to rule out the influence of response bias in

604   these results.

605

606   External load and measures of spatial perception (N=12): Seven studies used backpacks (all

607   approximately 20% of participants body weight), four used ankle weights (5-10% of body

608   weight), and one used resistance bands.

609   *Distance and gap estimation (N=11):* Pooling found that external loads resulted in participants

610   estimating distances as father than no load groups/conditions when using verbal (Figure 10a) and

611   VR based walking distance matching (Figure 10b) measures, but not blind walking measures

612   (Figure 10c). An additional study evaluated within-group differences in verbal estimates of

613   distance when wearing a backpack (Vinson et al. 2017), with findings consistent with pooled

614   results (unable to calculate effect size). The remaining experiments were also unable to be

615   pooled (Figure 10d). Of these, two (Corlett et al. 1990; White 2012) compared medium load (5kg

616   ankle weight/moderate resistance via band) conditions to high load (10km ankle weight/heavy

617   resistance via band) conditions. Both experiments had small, non-significant effects in the

618   direction that the high load group perceived distances as farther during blind walking measures

619   (Corlett et al. 1990) and VR based distance matching (White 2012). Last, one experiment found

620   that participants verbally estimated gap widths to be wider while wearing ankle weights than

621   they did without (Lessard et al. 2009). Sensitivity analyses evaluating the influence of varying

622   imputed correlation coefficients showed similar results (Supplementary File 5, Tables 8 and 9).

623   *Hill steepness estimations (N=1):* A single study evaluated the effect of external loads on verbal

624   estimation of hill steepness (Bhalla & Proffitt 1999), finding significant effects for a 5deg hill but

625   not a 31deg hill (Figure 10d). Opposite effects were seen for haptic measures of hill steepness.

626   *Interim discussion:* There was consistent evidence that external loads (versus no load) influence

627   verbal and VR-based measures of spatial perception in line with EoA, but conflicting evidence

628   for studies evaluating varying degrees of loading. Results require cautious interpretation given

629   experimental demands and a response bias. When two experiments (Keric & Sebanz 2021) used

630   identical protocols, but one used a cover story (successfully blinding participants) and one did

631   not (all participants guessed study hypothesis), the effect estimate was heightened in favour of

632   the unblinded group in line with EoA predictions (g= -0.32 vs -0.01 in the blinded group).

633   According to our risk of bias assessment, participants were not blinded in the remaining studies

634   in this grouping. Given that these unblinded studies have similar, large negative effects to the

635   unblinded Keric & Sebanz (2021) experiment, confidence in the overall pooled effect is reduced.

636   Further use of within-group, repeated measure design of some studies (e.g. (Lessard et al. 2009))

637   increase the possibility of response bias driving the effects seen.

638   Pooled results for VR experiments manipulating external load also showed a significant

639   effect in line with EoA predictions; however, all studies were from the same author (i.e., four

640   comparisons from two experiments), and are from a dissertation (not peer reviewed). Given

641   sample size reductions necessary for pooling, this resulted in small samples (n= 9-17), increasing

642   the risk of spurious results (Higgins et al. 2021). Additionally, assessor participation in measures

643   of spatial perception (e.g., assessor stopped the treadmill during VR-based walking distance

644   matching), paired with assessor non-blinding to test condition/study aim heightens the possibility

645   of measurement bias in this method.

646   Studies (Corlett et al. 1990; White 2012) investigating varying degrees of external loads

647   (high, low, no load) had non-significant results, but failure to assess participant-specific

648   perceptions of load magnitude hampers clear conclusions. It is unclear whether trends towards

649   increased distance estimation with high versus low loads represent the possibility of a dose-

650   response relationship, perhaps uncaptured if the loads applied were too light to induce a

651   homeostatic threat/influence (e.g., standard high load may not be perceived as high/difficult for

652   some people).

653

654   Effect of interoception (N=2): Two studies evaluated perceived internal bodily state (i.e.,

655   interoceptive accuracy) assessed using a heart rate accuracy test.

656   *Distance estimations (N=1):* The high interoceptive accuracy group (two-thirds of participants)

657   perceived the same distance as significantly farther on a hill than on the flat, in line with the EoA

658   (hills require higher energetic cost, therefore the same distance should be perceived as further

659   than on flat), but the low interoceptive group (remaining one-third of participants) did not

660   (Tenhundfeld & Witt 2015). Between-group differences were significant at only one of the four

661   distances (Hedge's g at 2m= 0.25 [95% CI – 0.35 to 0.85], 4m= -0.52 [95% CI -1.12 to 0.09],

662   6m= -0.64 [95% CI -1.25 to -0.03], and 8m= -0.6 [95% CI -1.21 to 0.01]).

663   *Hill steepness estimations (N=1):* A significant association was found between interoceptive

664   accuracy and verbal estimations of hill steepness (g=-0.89, 95% CI -1.36 to -0.41), indicating

665   that those with higher interoceptive accuracy had lower hill steepness estimations (over-

666   estimated steepness to a lesser degree) (Mouatt et al. 2021).

667   *Interim discussion:* Very limited evidence exists that an individual's ability to perceive their own

668   internal bodily state influences spatial perception. One study (Tenhundfeld & Witt 2015) had

669   high risk of bias (all domains assessed as high/unsure), and during correspondence, it was

670   reported that a replication study failed to demonstrate an effect (N. Tenhundfeld, personal

671   communication, 5th April, 2021). Thus, results here should be interpreted with caution.

672

673   ***Assessment of publication bias***: Contour enhanced funnel plots for each forest plot are given in

674   Supplementary File 6. Only the funnel plot for age met the suggested requirements of a

675   minimum of 10 data points to allow valid interpretation (Higgins et al. 2021). While funnel plot

676   is symmetrical around the summary effect size, several points are outside the p<0.01 boundary,

677   which may indicate publication bias. Due to significant concerns of publication bias in the field,

678   we created an additional post-hoc funnel plot including all studies with available data (Schäfer &

679   Schwarz 2019). In studies with multiple comparisons, we chose the comparison that was most

680   consistent with the overall result (if the majority comparisons were non-significant, we chose a

681   non-significant effect estimate; if the majority comparisons were significant, we chose a

682   significant effect estimate; if there was a mix of results, we chose an effect estimate at random).

683   Direction of effect was standardised such that a negative value indicates that a study's result

684 supports the EoA theory. One study was a clear outlier (effect size = 9.91); the funnel plot was

685 completed both with and without this study, with the latter allowing more robust visual

686 inspection of symmetry. Overall, there was adequate symmetry. However, 17% (n=9) of points

687 were outside the p<0.01 boundary in the direction of a negative effect, which may indicate

688 publication bias favouring the EoA hypothesis.

689

690 **Discussion:**

691   We found mixed support for the EoA hypothesis – that the environment would be

692 perceived as harsher (distances further, hills steeper) with reduced body capacity – across various

693 bodily states. The strongest and most consistent evidence in favour of the EoA was found for age

694 and glucose manipulation. There was mixed evidence for embodiment, body size manipulations,

695 fatigue, external loads, and interoception. Our risk of bias assessment indicates that theories of

696 social effects could only be ruled out in studies evaluating the effect of glucose due to their use

697 of adequate blinding procedures. There was no support for fitness influencing spatial perception

698 measures. However, lack of confirmation of successful induction of bodily states such as fatigue,

699 limit conclusions. That is, it remains unclear whether the tested bodily state does not influence

700 spatial perception or whether the tested bodily state was not actually successfully manipulated.

701 Moreover, given high risk of bias for all included studies, particularly in relation to response bias

702 and use of heterogenous spatial perception measures with unknown reliability, the existence of

703 an effect of any bodily state on spatial perception remains uncertain.

704   Our meta-analytical results are consistent with those of a previous review of more narrow

705 focus (Molto et al. 2020), which found evidence to support a small effect on visuospatial

706 perception of action-constraint (one's ability to act within the environment). This previous

707 review also investigated two commonly cited rival hypotheses, which both posit that the

708   observed effects for visuospatial perception are due to experimental demand biases, rather than

709   reflecting perceptual changes. The first hypothesis suggests that participants are more likely to

710   correctly guess the hypothesis in within-group designs than between group designs, and the

711   second that verbal measures of spatial perception may be more sensitive to voluntary control

712   than other measures (e.g., visual-matching or action-based). Thus, it is hypothesised that the

713   observed effects on spatial perception should be larger in studies using within-group designs and

714   verbal measures. Their moderator analysis did not support either hypothesis (Molto et al. 2020).

715   However, based on our findings of high overall risk of methodological bias of these studies, we

716   argue that such comparisons may be premature. That is, the null results of their moderator

717   analyses may reflect high variability of effects due to high risk of bias, rather than a true null

718   result.

719   Despite methodological issues of included studies, the current review revealed important

720   implications for the field of perceptual research. Across various bodily states, we found evidence

721   that perceptual scaling may require an interaction between bodily state and environmental

722   "difficulty". That is, perceptual scaling in line with the EoA may require both reduced bodily

723   capacity and an environment that is sufficiently "harsh" to induce homeostatic threat. For

724   instance, in studies evaluating fatigue, scaling occurred in hill steepness estimations but not in

725   distance estimations; this may represent insufficient "difficulty" of flat surfaces to evoke a threat

726   to homeostasis needed to scale visual perception when participants are likely not overly fatigued.

727   The results for age-induced effects on distance estimations also support this idea, where effect

728   sizes increased as the target distance (and thus homeostatic threat) increased. Homeostatic

729   specificity, and the interaction between bodily state and perceptual task, was also seen in studies

730   evaluating pain, where the results suggest that the measure of spatial perception must be relevant

731   to the site of pain to capture perceptual scaling (e.g., lower limb pain and walking distance

732   estimation, or shoulder pain and pointing distance estimation). Future work should ensure that

733   the environmental stimuli is both appropriately "difficult" and homeostatically relevant to induce

734   scaling given the target population. Exploring thresholds for difficulty and homeostatic relevance

735   on perception also merits consideration.

736   The interaction between bodily capacity and environmental features is also seen in cases

737   where perceptual scaling did *not* occur. In a study of Parkour athletes (Taylor et al. 2011),

738   estimations of wall heights were compared to an untrained control group, with a significant

739   perceptual benefit seen for athletes only for the shortest wall (1.94m) but not higher walls

740   (2.29m, 3.4m). At this low height, there may be a differential ability between groups to climb the

741   wall, enabling perceptual scaling, whereas for the higher walls, climbing was equally

742   unachievable for both groups, and perceptual scaling was absent. Similar results occurred for

743   distance estimations of gap width in weighted versus unweighted conditions (Lessard et al.

744   2009). A significant difference in distance estimation occurred, favouring smaller gap width

745   estimation for the unweighted condition, but disappeared when gaps were *outside* of participants'

746   estimated jumping ability (too far to jump across).

747   That similar bodily manipulations did not have uniform effects on perception, but rather,

748   were specific to the perceptual task provides support for EoA. For example, past work has shown

749   that embodying a small avatar (or the avatar of a child), increases estimated size of objects

750   (Banakou et al. 2013). Indeed, the findings of this review, where embodying a small avatar

751   increased perceived distance, extends this work, and suggests a similar perceptual scaling

752   mechanism, where larger bodies signify greater bodily capacity (van der Hoort et al. 2011). That

753   we view the world in context of our body is also supported by review findings for embodiment

754 showing that *having* a body in virtual space is a requirement for spatial scaling to occur (Phillips

755 et al. 2010; Ries et al. 2008; Scandola et al. 2019). Without a body, there is feasibly no internal

756 model that can inform capacity, and thus no alteration in environmental perception. If this latter

757 aspect is true, there are clear implications not only for the study of perception, but also in the

758 development of virtual reality applications.

759   Our findings also have important methodological implications for the broader field of

760 embodied perception. First, future research should ensure that experimental manipulations of

761 bodily state are achieved as intended. For example, when evaluating body size, assessing induced

762 alterations to perceived body size is needed to ensure successful manipulation. Additionally,

763 consideration and control of potential confounders is needed. The results of this meta-analysis

764 indicate that many factors relating to bodily state may influence spatial perception, and therefore

765 may act as confounders. For instance, if one was interested in the influence of pain on spatial

766 perception, one should ensure that the groups are matched by age to control for the confounding

767 influence of age on spatial perception. Or, when assessing the influence of fatigue on spatial

768 perception, it may also be relevant to consider body size, glucose levels, and/or fitness levels, as

769 these could influence perceptual scaling directly or via interactions with fatigue. Other factors,

770 such as affective state, may also be relevant to control for, however, it is beyond the scope of this

771 review to make such recommendations. Regardless, *a priori* consideration and pre-registration of

772 these variables is important for sample size calculations and to avoid type 1 error. Interestingly,

773 use of augmented or virtual reality to manipulate bodily state often significantly influenced

774 perception (Jun et al. 2015), while similar real-life manipulations did not (Collier 2018).

775 Determining if such conflicting results represent virtual environments not translating to real-life,

776 or insufficient/unbelievable real-life paradigms used is warranted.

777   Second, one consequence of the criticisms around measurement of spatial perceptual (i.e.,

778   that 'judgements' reflect response bias rather than perceptual effects, as previously highlighted)

779   is that indirect measures of spatial perception, such as action-based measures (blind walking,

780   throwing tasks) or visual matching paradigms, are favoured because they are thought more

781   robust against response bias. A presumed strength of these tasks is that participants use novel

782   motor skills (walking with a blindfold or throwing a bean bag), with task novelty reducing the

783   ability of participants to cognitively conform to perceived experimental demands. However, such

784   novelty has limitations. Motor learning, or the process through which one develops the ability to

785   perform a new action, is complex; accurately and consistently performing a motor task requires

786   time and practice (Krakauer et al. 2019). Thus, observed effects for these measures may reflect

787   poor and unreliable performances of the action-based task itself, rather than perceptual re-

788   scaling. For example, even if an individual perceived a distance to be near, a beanbag toss as the

789   measure of spatial perception ultimately relies upon their ability to avoid "overshooting" or

790   "undershooting" the target. Whether action-based tasks provide an accurate representation of

791   perceptual shifts is yet to be elucidated.

792   Challenges with reliability extends beyond action-based tasks. Our review found that only

793   four studies used spatial perception measures that had previously established reliability (See

794   Table 2). While the call for reliable tasks is not new – it was first discussed in 2013 and in many

795   subsequent papers (Firestone 2013; Philbeck & Witt 2015; Witt 2020) – the continued lack of

796   reliability assessment for spatial perception measures means that results are difficult to interpret.

797   For example, it is uncertain whether null results relate to a lack of precision of the spatial

798   measure or lack of perceptual effects. Ensuring reliability of spatial perception measures is a

799   priority for future research. Given our findings, we strongly recommend that expert consensus,

800   using methods such as a Delphi survey, to establish guidelines around methodological

801   considerations for reducing response bias and measuring spatial perception of the environment is

802   essential to move this field forward. Ideally, guideline development would involve experts from

803   many fields, including motor learning and blinding, to ensure that proposed strategies have

804   methodological rigour.

805   This review has several methodological strengths. Our protocol was developed *a priori* and

806   uploaded to Open Science Framework, with any deviations to the protocol disclosed and

807   reported. We used best practice methodology via compliance to the PRISMA 2021 statement

808   (Page et al. 2021). Two independent reviewers completed study screening, risk of bias

809   assessment, and data extraction. To minimise bias in the meta-analyses, comparisons chosen and

810   groupings for pooling were determined by an assessor blinded to the results of individual

811   experiments. The search strategy was comprehensive, and we contacted authors from 36 of 45

812   included studies to obtain missing data necessary for quantitative analysis.

813   Limitations of this review include the possibility of missing relevant studies given the lack of

814   established database subject headings for bodily state. However, this risk was mitigated using a

815   thorough systematic search strategy, informed by an academic librarian, including pearling of

816   reference lists, as well as consideration of grey literature. As with all meta-analyses, there is risk

817   that effect sizes may be artificially inflated as this method can amplify publication bias (Friese &

818   Frankenbach 2020). Within psychology, there is a general consensus that publication bias is

819   present, however the exact prevalence is unknown (Schäfer & Schwarz 2019). Therefore, it is

820   difficult to estimate the degree to which these factors influenced the results of this study. We

821   attempted to mitigate the effects of publication bias by including grey literature, however there

822   may still be systemic error present in our results. Additionally, a possible limitation exists via the

823 graphical extraction of outcome data using electronic tools in experiments where numeric data

824 were not available. However, our chosen method has good reliability (Drevon et al. 2017), and

825 any errors occurring from extraction are unlikely to be systematic. Last, it is possible that our

826 effect estimates for within-group studies were influenced by imputation of the correlation

827 between the pair of observations (r). We were able to calculate the value of r from raw data in 23

828 comparisons, but for the remaining we imputed r using values from similar experiments, as

829 recommended (Higgins et al. 2021). Importantly, that our sensitivity analyses (increasing and

830 decreasing r by 0.1) did not change our results, suggests that our findings are robust.

831

832 **Conclusion:**

833 This meta-analytical systematic review found low-level evidence for the effect of bodily state on

834 measures of spatial perception, given the high risk of bias of eligible studies. However,

835 consistent evidence exists to support that glucose manipulations and age influence spatial

836 perception of distance and hill steepness in line with the Economy of Action hypothesis where

837 lower capacity results in over-estimation of distance/steepness (and vice versa). In the glucose

838 grouping, adequate blinding and study design reduces the likelihood of competing theories

839 (social and knowledge effects) explaining the results, thus increasing the confidence that the

840 results reflect true shifts in spatial perception. Methodological limitations of this field, including

841 the lack of confirmed success of bodily state manipulations and the lack of consideration of

842 relevant confounders, make interpretation of both significant and null results challenging.

843 Establishing the reliability of spatial perception measures and employing experimental methods

844 to reduce participant response bias (including blinding) are priorities for future research.

846   **References**

847   Alaiti RK, Harvie DS, Gasparin JT, de Sousa MF, Pompeu JE, Madden VJ, Leite Hunziker MH,
848   and da Costa MF. 2019. Distorted distance perception to reachable points in people
849   with chronic shoulder pain. *Musculoskelet Sci Pract* 42:120-124.
850   10.1016/j.msksp.2019.04.015
851   Asaf DA, Santillán JE, and Barraza JF. 2015. Visuo-Motor Interaction in the Estimation of
852   Distance: Athletes vs. No-Athletes. Cham: Springer International Publishing. p 265-
853   268.
854   Baati H, Chtourou H, Moalla W, Jarraya M, Nikolaidis PT, Rosemann T, and Knechtle B.
855   2020. Effect of Angle of View and Partial Sleep Deprivation on Distance Perception.
856   *Front Psychol* 11:201-201. 10.3389/fpsyg.2020.00201
857   Baati H, Shell Hmani M, Jarraya M, Chtourou H, Masmoudi L, Trabelsi K, Hakim A,
858   Marrakchi R, and Moalla W. 2015. Effect of total sleep deprivation on egocentric
859   distance estimation following a fatiguing task. *Biological rhythm research* 46:265-
860   274. 10.1080/09291016.2014.985003
861   Balduzzi S, Rücker G, and Schwarzer G. 2019. How to perform a meta-analysis with R: a
862   practical tutorial. *Evidence Based Mental Health* 22:153. 10.1136/ebmental-2019-
863   300117
864   Banakou D, Groten R, and Slater M. 2013. Illusory ownership of a virtual child body causes
865   overestimation of object sizes and implicit attitude changes. *Proceedings of the*
866   *National Academy of Sciences* 110:12846-12851.
867   Bhalla M, and Proffitt D. 1999. Visual-motor recalibration in geographical slant perception.
868   *Journal of Experimental Psychology* 25:1076-1096. 10.1037/0096-1523.25.4.1076
869   Bian Z, and Andersen GJ. 2013. Aging and the perception of egocentric distance. *Psychology*
870   *and aging* 28:813.
871   Borenstein M, Hedges LV, Higgins JPT, and Rothstein HR. 2011. *Introduction to Meta-*
872   *Analysis*. Somerset: Somerset: Wiley.
873   Bridgeman B, and Cooke I. 2015. Effect of eye height on estimated slopes of hills. *Perception*
874   44:755-763.
875   Clark A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive
876   science. *Behav Brain Sci* 36:181-204. 10.1017/S0140525X12000477
877   Clark A. 2018. A nice surprise? Predictive processing and the active pursuit of novelty.
878   *Phenomenology and the cognitive sciences* 17:521-534. 10.1007/s11097-017-9525-z
879   Cole S, and Balcetis E. 2013. Sources of Resources: Bioenergetic and Psychoenergetic
880   Resources Influence Distance Perception. *Social cognition* 31:721-732.
881   10.1521/soco.2013.31.6.721
882   Cole S, Balcetis E, and Zhang S. 2013. Visual perception and regulatory conflict: Motivation
883   and physiology influence distance perception. *Journal of Experimental Psychology:*
884   *General* 142:18.
885   Collier ES. 2018. The illusion of action-specific scaling effects: action capacity does not
886   directly influence spatial perception. ProQuest Dissertations Publishing.
887   Corlett JT, Byblow W, and Taylor B. 1990. The Effect of Perceived Locomotor Constraints on
888   Distance Estimation. *Journal of Motor Behavior* 22:347-360.
889   10.1080/00222895.1990.10735518

890   Costello MC, Bloesch EK, Davoli CC, Panting ND, Abrams RA, and Brockmole JR. 2015.
891   Spatial Representations in Older Adults Are Not Modified by Action: Evidence From
892   Tool Use. *Psychol Aging* 30:656-668. 10.1037/pag0000029
893   Dean AM, Oh J, Thomson CJ, Norris CJ, and Durgin FH. 2016. Do Individual Differences and
894   Aging Effects in the Estimation of Geographical Slant Reflect Cognitive or Perceptual
895   Effects? *Iperception* 7:1-18. 10.1177/2041669516658665
896   Downes MJ, Brennan ML, Williams HC, and Dean RS. 2016. Development of a critical
897   appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ open*
898   6:e011458.
899   Drevon D, Fursa SR, and Malcolm AL. 2017. Intercoder reliability and validity of
900   WebPlotDigitizer in extracting graphed data. *Behavior modification* 41:323-339.
901   Durgin FH, Baird JA, Greenburg M, Russell R, Shaughnessy K, and Waymouth S. 2009. Who
902   is being deceived? The experimental demands of wearing a backpack. *Psychon Bull*
903   *Rev* 16:964-969. 10.3758/pbr.16.5.964
904   Durgin FH, Klein B, Spiegel A, Strawser CJ, and Williams M. 2012. The social psychology of
905   perception experiments: hills, backpacks, glucose, and the problem of
906   generalizability. *Journal of Experimental Psychology: Human Perception and*
907   *Performance* 38:1582.
908   Egger M, Smith GD, Schneider M, and Minder C. 1997. Bias in meta-analysis detected by a
909   simple, graphical test. *BMJ* 315:629-634. 10.1136/bmj.315.7109.629
910   Firestone C. 2013. How "Paternalistic" Is Spatial Perception? Why Wearing a Heavy
911   Backpack Doesn't—and Couldn't —Make Hills Look Steeper. *Perspect Psychol Sci*
912   8:455-473. 10.1177/1745691613489835
913   Firestone C, and Scholl BJ. 2017. Seeing and Thinking in Studies of Embodied "Perception":
914   How (Not) to Integrate Vision Science and Social Psychology. *Perspect Psychol Sci*
915   12:341-343. 10.1177/1745691616679944
916   Friese M, and Frankenbach J. 2020. p-Hacking and publication bias interact to distort meta-
917   analytic effect size estimates. *Psychological Methods* 25:456.
918   Friston K. 2010. The free-energy principle: a unified brain theory? *Nature Reviews*
919   *Neuroscience* 11:127-138. 10.1038/nrn2787
920   Friston K. 2011. Embodied inference: or" I think therefore I am, if I am what I think".
921   Hedges LV. 1981. Distribution theory for Glass's estimator of effect size and related
922   estimators. *journal of Educational Statistics* 6:107-128.
923   Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, and Welch V. 2021. Cochrane
924   Handbook for Systematic Reviews of Interventions version 6.2 (updated February
925   2021). Cochrane.
926   Hunt DF, Hunt HBN, and Park JH. 2017. Bioenergetic costs and state influence distance
927   perception. *Physiol Behav* 180:103-106. 10.1016/j.physbeh.2017.08.003
928   Hutchison JJ, and Loomis JM. 2014. Does Energy Expenditure Affect the Perception of
929   Egocentric Distance? A Failure to Replicate Experiment 1 of Proffitt, Stefanucci,
930   Banton, and Epstein (2003). *The Spanish journal of psychology* 9:332-339.
931   10.1017/s1138741600006235
932   Jarraya M, Chtourou H, Megdich K, Chaouachi A, Souissi N, and Chamari K. 2013. Effect of a
933   Moderate-Intensity Aerobic Exercise on Estimates of Egocentric Distance. *Percept*
934   *Mot Skills* 116:658-670. 10.2466/06.24.PMS.116.2.658-670

935 Jun E, Stefanucci J, Creem-Regehr S, Geuss M, and Thompson W. 2015. Big Foot: Using the
936   Size of a Virtual Foot to Scale Gap Width. *ACM transactions on applied perception*
937   12:1-12. 10.1145/2811266

938 Keric V, and Sebanz N. 2021. Task Construal Influences Estimations of the Environment.
939   *Frontiers in Human Neuroscience*.

940 Krakauer JW, Hadjiosif AM, Xu J, Wong AL, and Haith AM. 2019. Motor learning. *Compr*
941   *Physiol* 9:613-663.

942 Krpan D, and Schnall S. 2014. When perception says "no" to action: Approach cues make
943   steep hills appear even steeper. *Journal of Experimental Social Psychology* 55:89-98.
944   10.1016/j.jesp.2014.06.005

945 Lessard DA, Linkenauger SA, and Proffitt DR. 2009. Look before You Leap: Jumping Ability
946   Affects Distance Perception. *Perception* 38:1863-1866. 10.1068/p6509

947 McAuley L, Pham B, Tugwell P, and Moher D. 2000. Does the inclusion of grey literature
948   influence estimates of intervention effectiveness reported in meta-analyses? *The*
949   *Lancet* 356:1228-1231. https://doi.org/10.1016/S0140-6736(00)02786-0

950 Molto L, Nalborczyk L, Palluel-Germain R, and Morgado N. 2020. Action Effects on Visual
951   Perception of Distances: A Multilevel Bayesian Meta-Analysis. *Psychol Sci* 31:488-
952   504. 10.1177/0956797619900336

953 Mouatt B, Smith A, Parfitt G, Stanford T, McDade J, Smith RT, and Stanton TR. 2021. What I
954   see and what I feel: The influence of exteroceptive and interoceptive cues on
955   affective valence and sense of effort during exercise. *[Manuscript submitted for*
956   *publication]*.

957 Nevill AM, Stewart AD, Olds T, and Holder R. 2006. Relationship between adiposity and
958   body size reveals limitations of BMI. *American Journal of Physical Anthropology: The*
959   *Official Publication of the American Association of Physical Anthropologists* 129:151-
960   156.

961 Norman JF, Dukes JM, Shapiro HK, and Peterson AE. 2020. The Visual Perception of Large-
962   Scale Distances Outdoors. *Perception* 49:968-977.

963 Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L,
964   Tetzlaff JM, Akl EA, and Brennan SE. 2021. The PRISMA 2020 statement: an updated
965   guideline for reporting systematic reviews. *BMJ* 372.

966 Philbeck JW, and Witt JK. 2015. Action-specific influences on perception and
967   postperceptual processes: Present controversies and future directions. *Psychol Bull*
968   141:1120-1144. 10.1037/a0039738

969 Phillips L, Ries B, Kaeding M, and Interrante V. 2010. Avatar self-embodiment enhances
970   distance perception accuracy in non-photorealistic immersive virtual environments.
971   2010 IEEE virtual reality conference (VR): IEEE. p 115-1148.

972 Proffitt DR. 2006. Embodied Perception and the Economy of Action. *Perspect Psychol Sci*
973   1:110-122. 10.1111/j.1745-6916.2006.00008.x

974 Proffitt DR, Bhalla M, Gossweiler R, and Midgett J. 1995. Perceiving geographical slant.
975   *Psychonomic Bulletin & Review* 2:409-428.

976 Ries B, Interrante V, Kaeding M, and Anderson L. 2008. The effect of self-embodiment on
977   distance perception in immersive virtual environments. Proceedings of the 2008
978   ACM symposium on Virtual reality software and technology. p 167-170.

979 Rohatgi A. 2020. WebPlotDigitizer. *Available at*
980   *https://automeris.io/WebPlotDigitizer*2021).

981 Roth PL, Le H, Oh I-S, Van Iddekinge CH, and Bobko P. 2018. Using beta coefficients to
982   impute missing correlations in meta-analysis research: Reasons for caution. *Journal*
983   *of Applied Psychology* 103:644.
984 Scandola M, Togni R, Tieri G, Avesani R, Brambilla M, Aglioti SM, and Moro V. 2019.
985   Embodying their own wheelchair modifies extrapersonal space perception in people
986   with spinal cord injury. *Exp Brain Res* 237:2621-2632. 10.1007/s00221-019-05618-
987   8
988 Schäfer T, and Schwarz MA. 2019. The Meaningfulness of Effect Sizes in Psychological
989   Research: Differences Between Sub-Disciplines and the Impact of Potential Biases.
990   *Frontiers in psychology* 10. 10.3389/fpsyg.2019.00813
991 Schnall S, Zadra JR, and Proffitt DR. 2010. Direct Evidence for the Economy of Action:
992   Glucose and the Perception of Geographical Slant. *Perception (London)* 39:464-482.
993   10.1068/p6445
994 Shaffer DM, Greer KM, and Schaffer JT. 2019. Manipulation of Expended Effort and Intent
995   Does Not Affect Estimates of Slant or Distance. *J Exp Psychol Hum Percept Perform*
996   45:855-862. 10.1037/xhp0000641
997 Shaffer DM, McManama E, Swank C, and Durgin FH. 2013. Sugar and space? Not the case:
998   Effects of low blood glucose on slant estimation are mediated by beliefs. *i-Perception*
999   4:147-155.
1000 Sugovic M, Turk P, and Witt JK. 2016. Perceived distance and obesity: It's what you weigh,
1001   not what you think. *Acta psychologica* 165:1-8. 10.1016/j.actpsy.2016.01.012
1002 Sugovic M, and Witt JK. 2013. An older view on distance perception: Older adults perceive
1003   walkable extents as farther. *Experimental Brain Research* 226:383-391.
1004 Tabor A, O'Daly O, Gregory RW, Jacobs C, Travers W, Thacker MA, and Moseley GL. 2016.
1005   Perceptual Inference in Chronic Pain: An Investigation Into the Economy of Action
1006   Hypothesis. *Clin J Pain* 32:588-593. 10.1097/AJP.0000000000000305
1007 Taylor JET, Witt JK, and Sugovic M. 2011. When Walls are No Longer Barriers: Perception of
1008   Wall Height in Parkour. *Perception* 40:757-760. 10.1068/p6855
1009 Taylor-Covill GAH, and Eves FF. 2013. Slant Perception for Stairs and Screens: Effects of Sex
1010   and Fatigue in a Laboratory Environment. *Perception* 42:459-469. 10.1068/p7425
1011 Taylor-Covill GAH, and Eves FF. 2016. Carrying a biological "backpack": Quasi-experimental
1012   effects of weight status and body fat change on perceived steepness. *J Exp Psychol*
1013   *Hum Percept Perform* 42:331-338. 10.1037/xhp0000137
1014 Team RC. 2021. R: A language and environment for statistical computing. *Available at*
1015   *https://www.R-project.org/*.
1016 Tenhundfeld NL, and Witt JK. 2015. The effects of interoceptive attunement on action-
1017   specific perception. *Visual cognition* 23:837-840. 10.1080/13506285.2015.1093240
1018 Thayer RE. 1978. Toward a psychological theory of multidimensional activation (arousal).
1019   *Motivation and emotion* 2:1-34. 10.1007/BF00992729
1020 van der Hoort B, Guterstam A, and Ehrsson HH. 2011. Being Barbie: The Size of One's Own
1021   Body Determines the Perceived Size of the World. *PLoS ONE* 6:e20195-e20195.
1022   10.1371/journal.pone.0020195
1023 Viechtbauer W. 2010. Conducting meta-analyses in R with the metafor package. *Journal of*
1024   *statistical software* 36:1-48.

1025   Vinson DW, Vinson DW, Jordan JS, Jordan JS, Hund AM, and Hund AM. 2017. Perceptually
1026   walking in another's shoes: goals and memories constrain spatial perception.
1027   *Psychol Res* 81:66-74. 10.1007/s00426-015-0714-5
1028   White E. 2012. The role of multimodally specified effort in action-relevant distance
1029   perception. University of Cincinnati / OhioLINK.
1030   Wier LT, Jackson AS, Ayers GW, and Arenare B. 2006. Nonexercise models for estimating
1031   VO2max with waist girth, percent fat, or BMI. *Medicine and science in sports and*
1032   *exercise* 38:555-561.
1033   Witt J, Linkenauger S, Bakdash J, Augustyn J, Cook A, and Proffitt D. 2009. The long road of
1034   pain: chronic pain increases perceived distance. *Experimental Brain Research*
1035   192:145-148. 10.1007/s00221-008-1594-3
1036   Witt JK. 2020. Action's influence on spatial perception: resolution and a mystery. *Current*
1037   *Opinion in Psychology* 32:153-157. https://doi.org/10.1016/j.copsyc.2019.07.045
1038   Zadra JR. 2013. Bioenergetic Effects on Perception and Cognition. University of Virginia.
1039   Zadra JR, Weltman AL, and Proffitt DR. 2016. Walkable distances are bioenergetically
1040   scaled. *J Exp Psychol Hum Percept Perform* 42:39-51. 10.1037/xhp0000107
1041

Figure 1

PRISMA flowchart detailing the search, screening and inclusion process

Identification of new studies via databases

Identification of new studies via other methods

Identification

Records identified from:
Databases (n =8034)
Medline (OVID) (n=3264)
PsychINFO (OVID) (n= 1666)
Embase (OVID) (n=1377)
Emcare (OVID) (n=399)
Scopus (Elsevier) (n=1328)

→ Records removed before screening:
Duplicate records: (n =2536)

Records identified from:
Conference abstracts (n= 15)
ProQuest Theses and Dissertations (n=74)
Google Scholar (n=93)
Pearling (n= 17)
Web of Science Citation Tracking (n=9)

Screening

Records screened
(n =5498)

→ Records excluded
(n =5402)

Records screened

→ Records excluded
(n =186)

Reports sought for retrieval
(n =96)

→ Reports not retrieved
(n =2)

Reports sought for retrieval
(n =25)

→ Reports not retrieved
(n = 1)

Reports assessed for eligibility (n =94), containing 179 experiments

→ Reports excluded:
(n=132 experiments)
Due to measure of spatial perception (n=77)
Due to measure of bodily state (n= 43)
Due to study design (n=1)
Duplicate (n=8)
Other (n=3)

Reports assessed for eligibility (n = 24) containing 55 experiments

→ Experiments excluded:
(n=35)
Due to measure of spatial perception (n=24)
Due to measure of bodily state (n=7)
Duplicate (n=3)
Other (n=1)

Included

Experiments from databases included (n = 47)
Experiments from grey literature included (n =21)

Total experiments included in review (n =68)

Figure 2

Scores of each category in the Risk of Bias assessment, illustrated by percentage

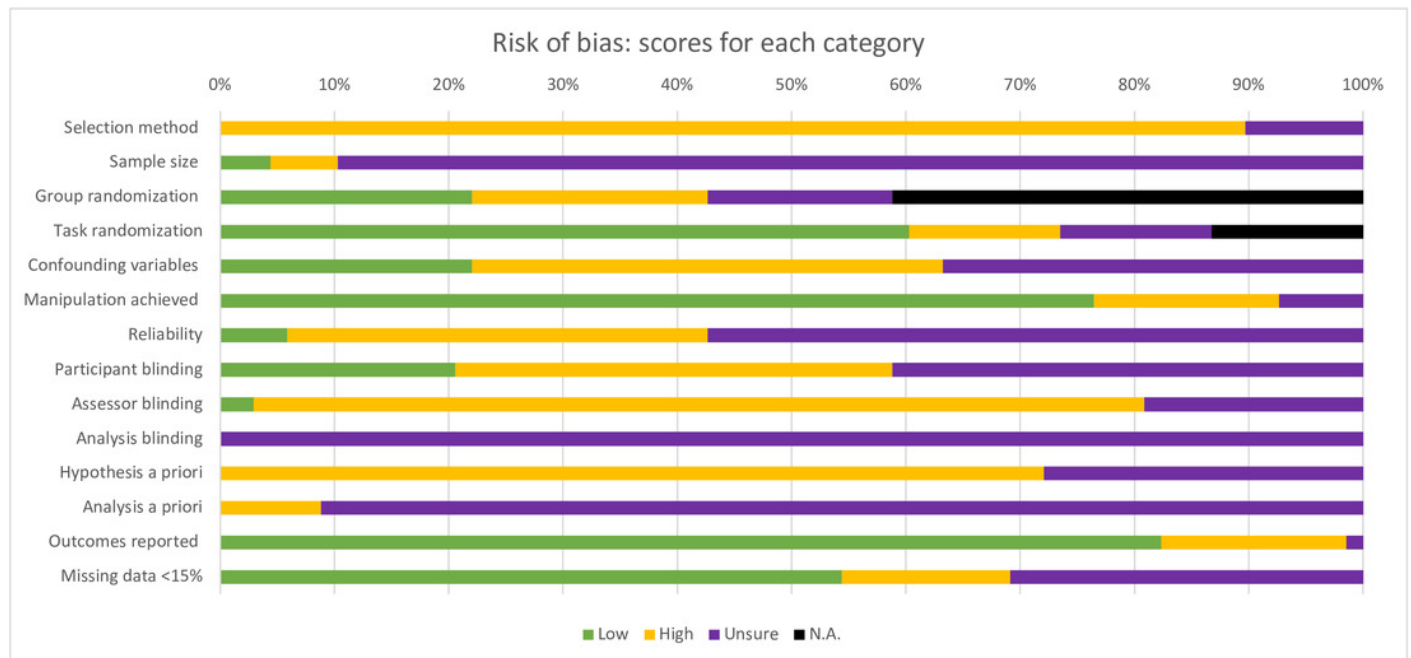



Figure 3

Pooled estimates for the effect of fatigue on spatial perception.

(A) Verbal distance estimation pre- and post-exercise, subgrouped by distance. (B) Verbal distance estimation preand post-exercise during conditions of sleep deprivation, subgrouped by distance. (C) Verbal distance estimation in conditions of normal sleep and sleep deprivation, subgrouped by distance. (D) Verbal hill steepness estimations pre- and post-exercise, subgrouped by hill steepness. (E) Haptic estimations of hill steepness pre- and post-exercise, subgrouped by hill steepness. In all cases a negative effect indicates that the fatigued group estimated spatial perception measures (distance or steepness) as larger than the control group.

a

b

c

d

e

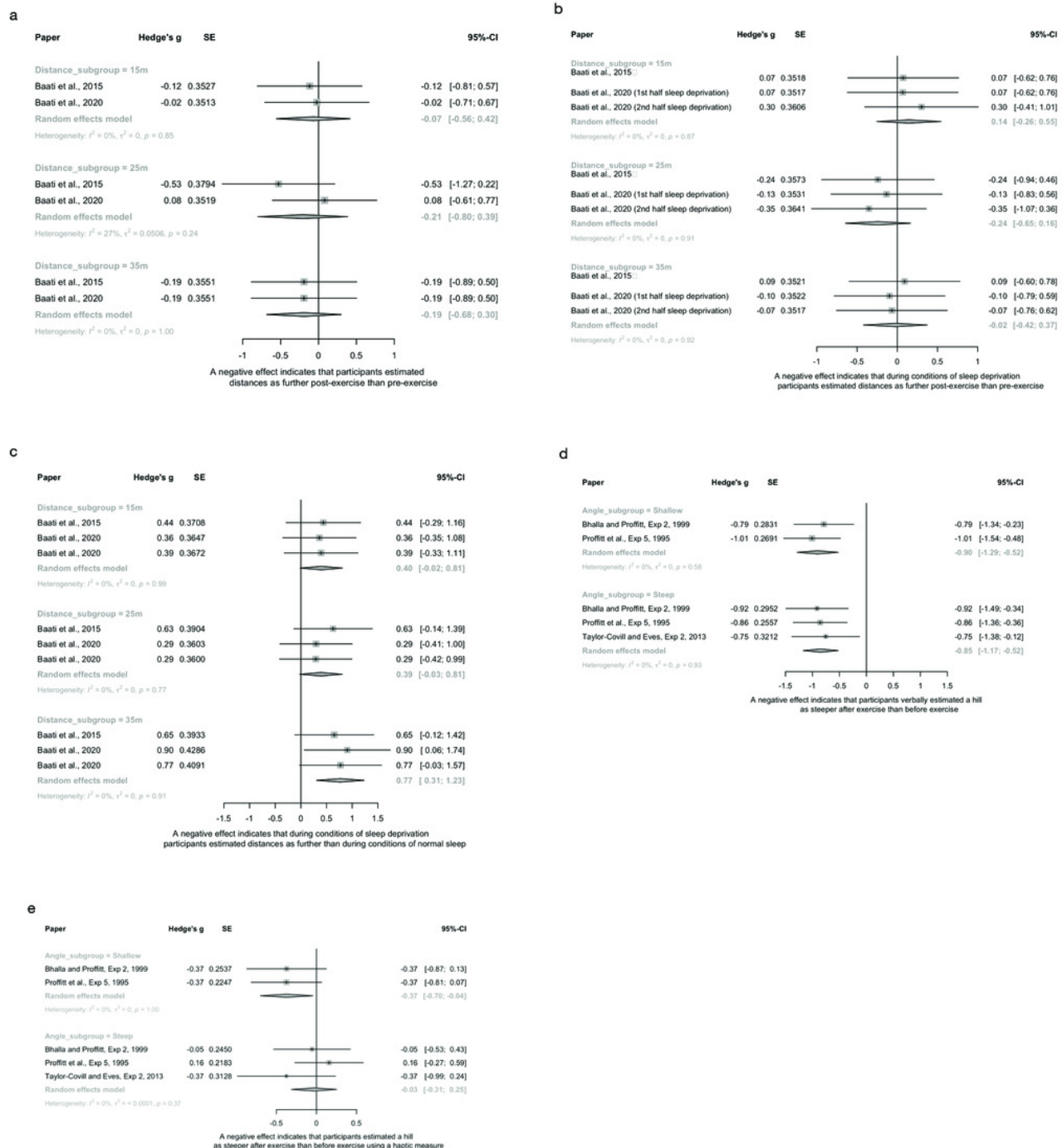



Figure 4

Pooled estimates for the effect of pain on spatial perception.

A negative effect means that the chronic pain group estimated distances as farther than the control group.

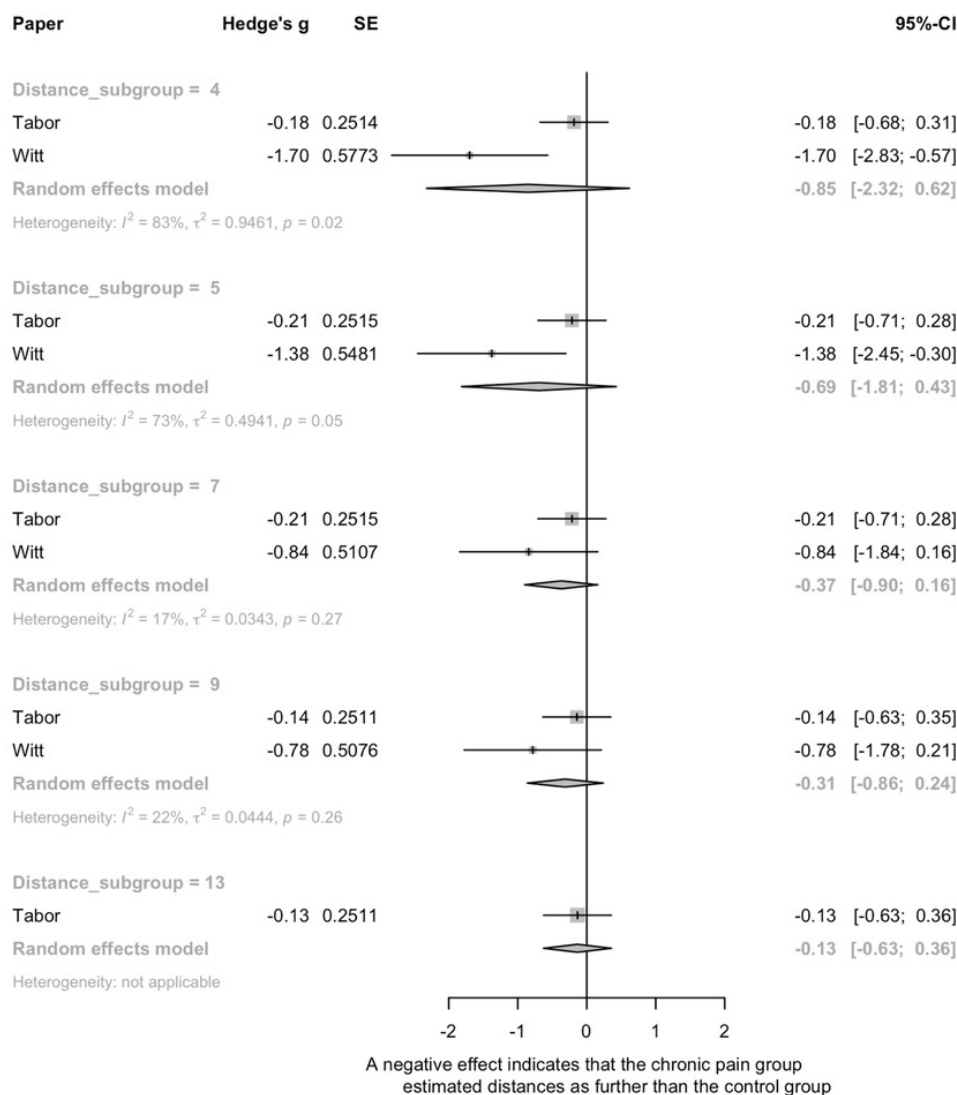



Figure 5

Pooled estimates for the effect of age on measures of spatial perception.

(A) Verbal distance estimation, subgrouped by distance. (B) Individual study effect sizes. Error bars are 95% CIs and * indicates a significant effect. (C) Effect sizes of all hills from Bhalla 1999, sorted by type of spatial perception measure used. Error bars are 95% CIs and * indicates a significant effect. In all cases, a negative effect indicates that the older group estimated distances as farther than the younger group.

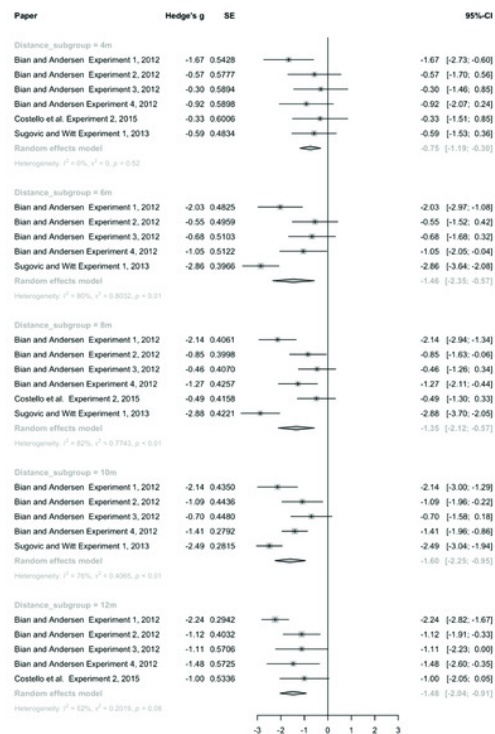
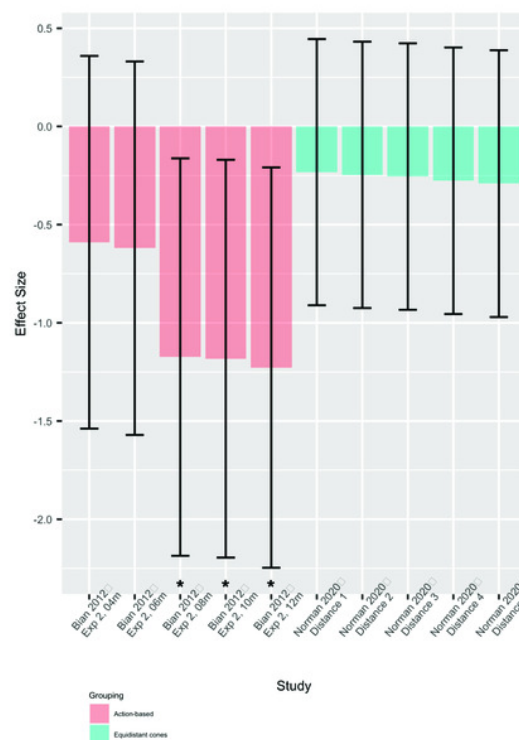
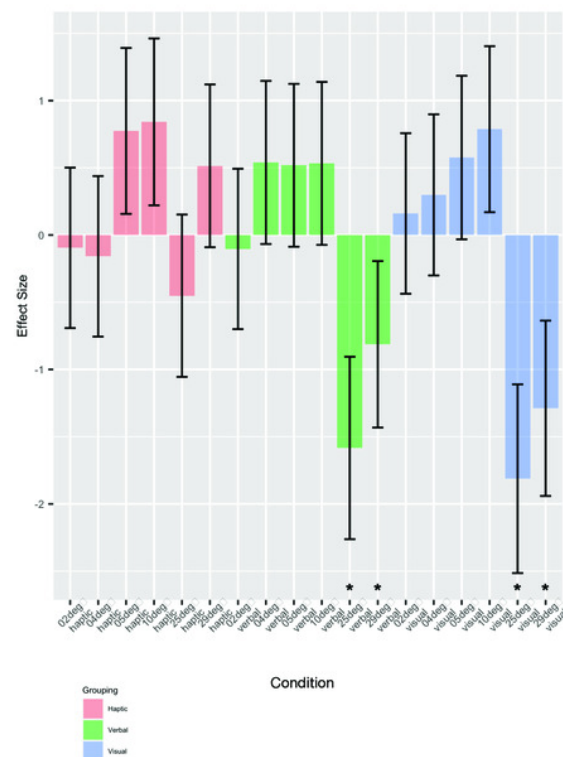



Figure 6

The effect of body size manipulations on measures of spatial perception.

(A) Pooled estimates for the effect of a hand size manipulation on a visual-matching aperture width task. (B) Individual study effect sizes, sorted by experimental manipulation. Error bars are 95% CIs and * indicates a significant effect. (C) Pooled estimates for the effect of height manipulation on verbal steepness estimations. In all cases, positive effect indicates that the larger body/body part size group estimated distances as closer/hills as less steep than the smaller body/body part size group

a

Paper	Hedge's g	SE		95%-CI	Weight
Collier Ch 5, Exp 1	0.00	0.2567		0.00 [-0.50; 0.51]	47.2%
Collier Ch 5, Exp 2	0.04	0.3434		0.04 [-0.63; 0.71]	26.4%
Collier Ch 5, Exp 2	0.01	0.3433		0.01 [-0.66; 0.68]	26.4%
Random effects model				**0.01 [-0.33; 0.36]**	**100.0%**

Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0$, $p = 1.00$

A positive effect indicates that apperatures were estimated as
larger in the unpadded (normal hand) group than the padded (large hand) group

b

Study

Grouping
- Blind-walking
- Verbal

c

A positive effect indicates that participants standing
on a box percieved the hill as less steep than those standing on the ground

Figure 7

The effect of embodiment on measures of spatial perception.

(A) Pooled estimates for the effect of embodiment on blind walking distance estimation. (B) Individual effect sizes of embodiment (own vs. other wheelchair) on verbal estimates of distance on hill estimations. (C) Individual effect sizes of embodiment (own vs. other wheelchair) on a visual matching hill steepness estimations. Of note, prior to calculating effect sizes from author supplied data for hill steepness estimates, 8.9% (n= 160) were removed from the dataset, as the estimations were implausible (>90deg, where 90 degrees is a vertical wall). In all cases, a negative effect indicates that the embodied group estimated distances as farther than the not embodied group.

a

Paper	Hedge's g	SE		95%-CI	Weight
Phillips et al., 2010	-0.94	0.4536		-0.94 [-1.83; -0.06]	49.5%
Ries et al., 2008	-0.86	0.4495		-0.86 [-1.74; 0.02]	50.5%
Random effects model				**-0.90 [-1.53; -0.28]**	**100.0%**

Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0$, $p = 0.90$

A positive effect indicates that the avatar group estimated
distances as closer than the no-avatar group

b

c

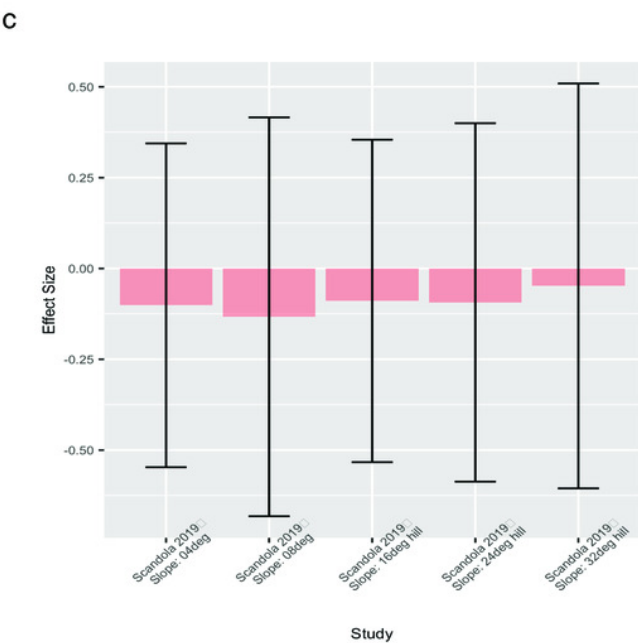
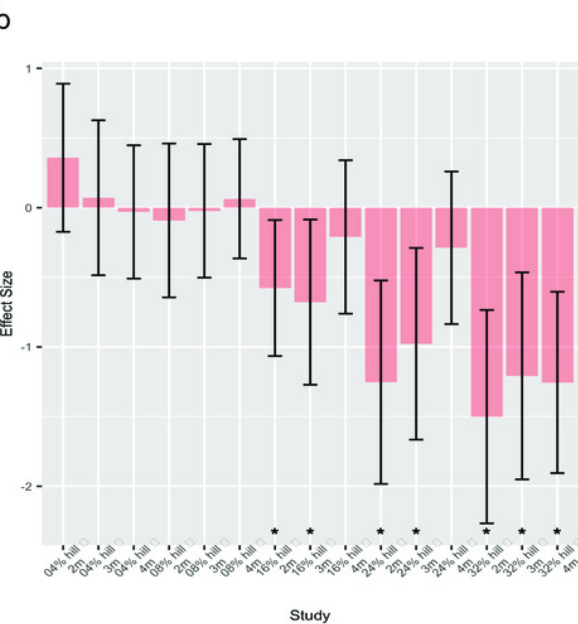
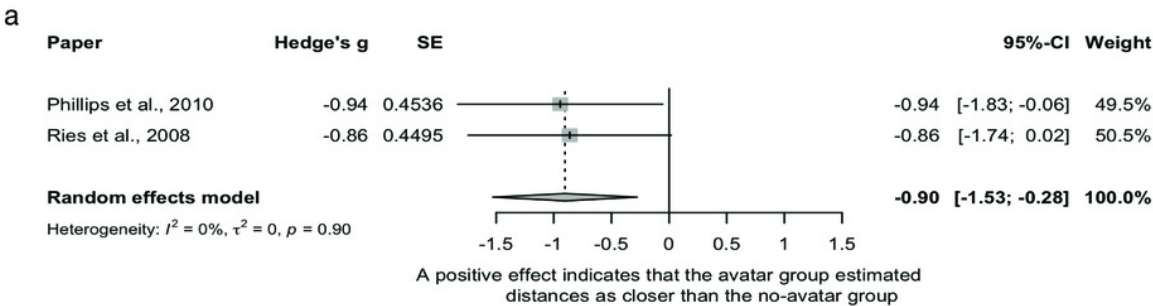
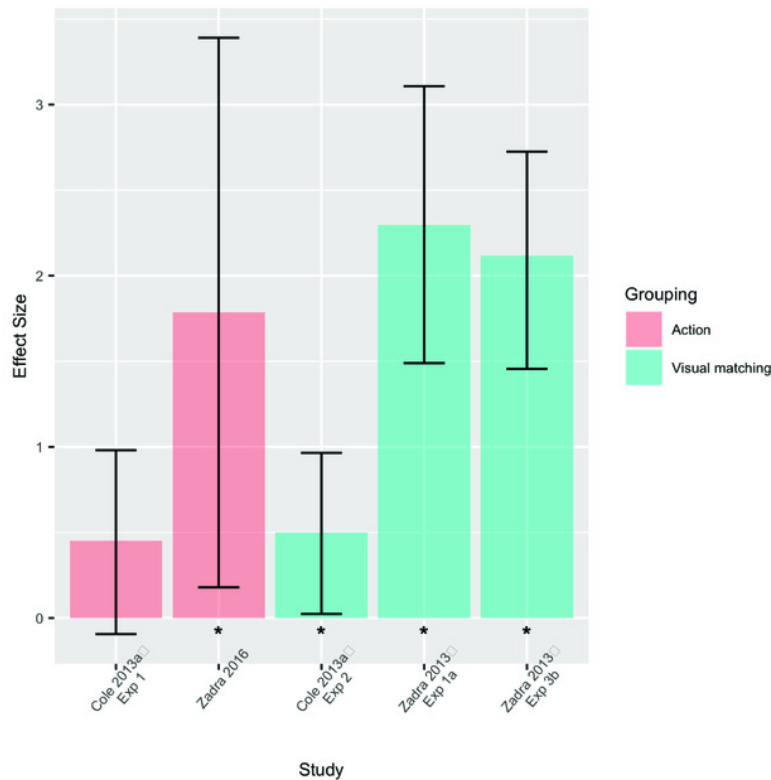



Figure 8

The effect of glucose manipulations on measures of spatial perception.

(A) Individual study effect sizes, sorted by type of spatial perception measure. Error bars are 95% CIs and * represents significant effects. (B) Pooled verbal estimations of the effect of glucose manipulations on verbal hill steepness, subgrouped by steepness.

a

b

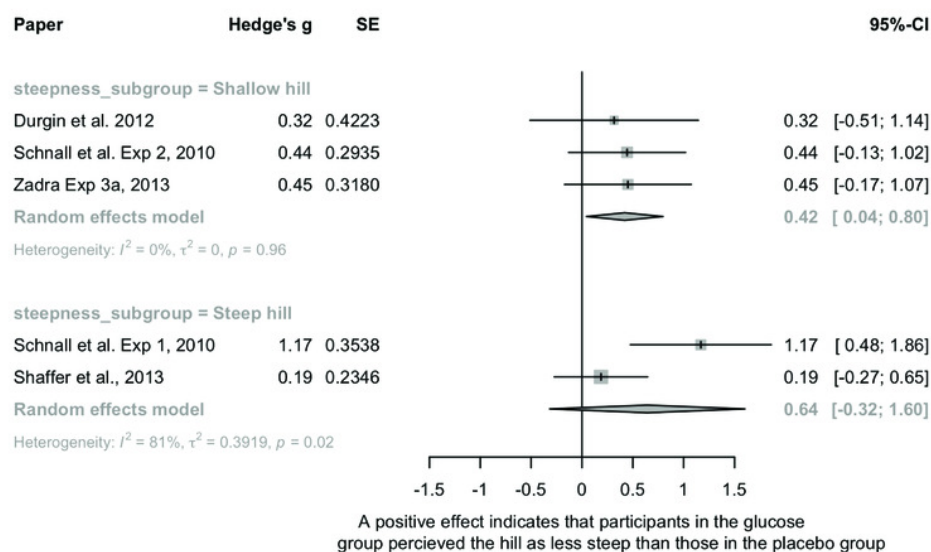
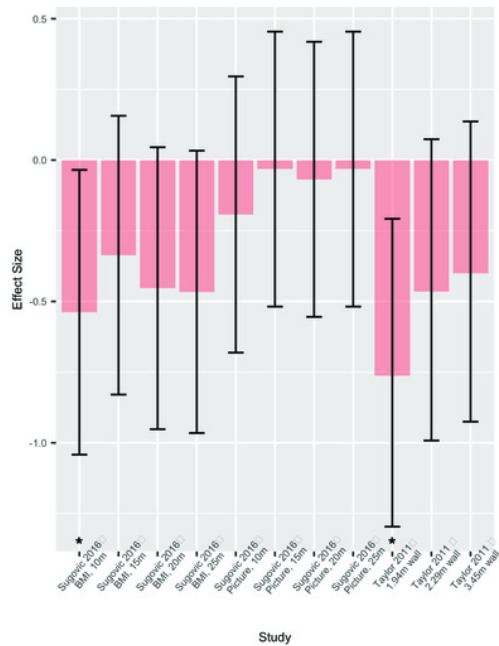
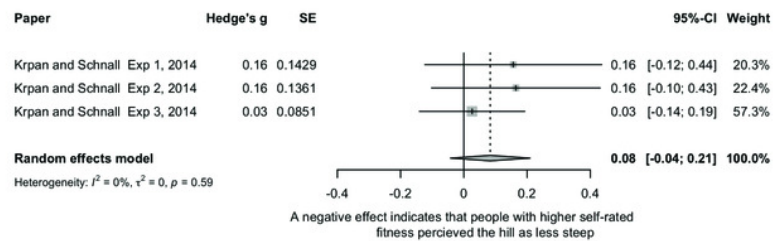
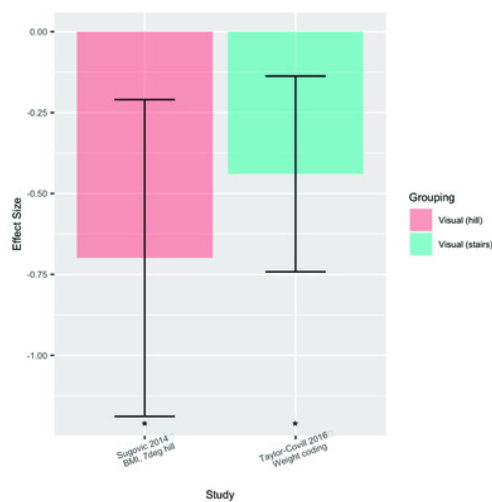



Figure 9

The effect of fitness loads on measures of spatial perception.

(A) Individual distance estimation experiment effect sizes. Error bars are 95% CIs and *
indicates significant effects. (B) Pooled estimates for self rated fitness and verbal hill
estimation. (C) Pooled estimates for self-rated fitness and haptic hill estimations. (D)
Individual steepness estimation experiment effect sizes. Error bars are 95% CIs and *
indicates significant effects. In all cases a positive effect indicates that the higher fitness
group estimated measures of spatial perception (distances and steepness) as larger than the
lower fitness group.

a

b

d

c

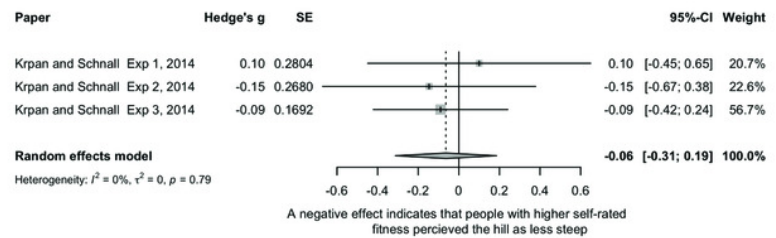



Figure 10

The effect of external loads on measures of spatial perception.

(A) Pooled estimates of the effect of external loads on verbal estimation of distance. (B) Pooled estimates of the effect of external loads on a VR matching distance paradigm. (C) Pooled estimates of the effect of external loads on blind0-walking distance estimation. (D) Individual study effect sizes. Error bars are 95% CIs and * indicates significant effects. In all cases a negative effect indicates that the external load group estimated measures of spatial perception (distances and steepness) as larger than the control group.

a

b

c

d

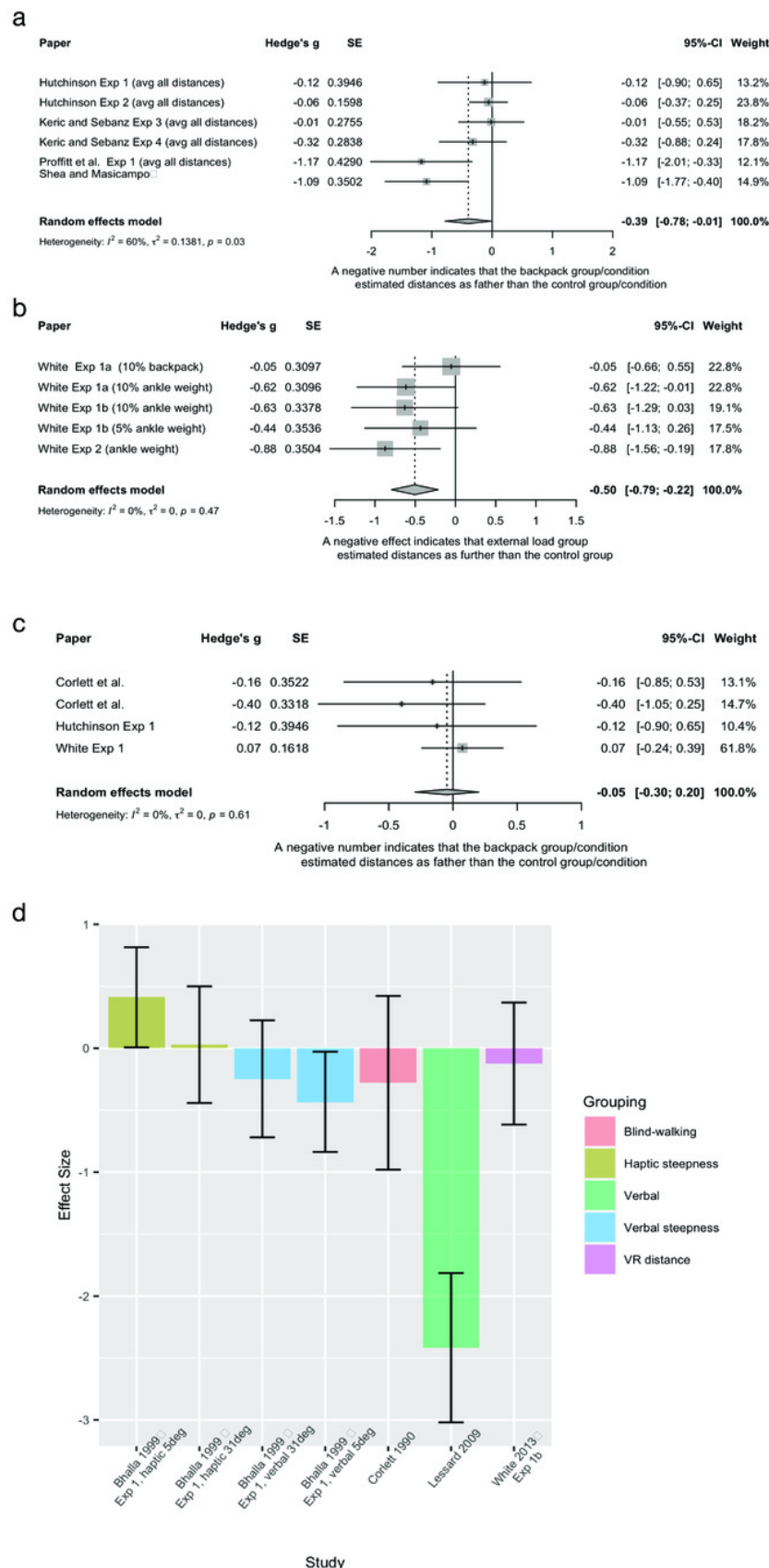



Table 1: Characteristics of Included Studies

Grouping	Study Details		Study Design	Participant Details	Bodily State Comparison, Manipulation, and/or Association	Spatial Perception Outcome
Age	Bhalla and Proffitt Experiment 4	1999	Between group, cross sectional observational study	Older adults group n= 32, sex not reported; mean age 73 years, age range from 60-87 years (all >60 years of age) "Student aged" normative dataset n=300, 150(f), age not reported The "student aged" group was previously collected as part of a normative dataset (Proffitt et al., 1995)	Older adults vs "student aged" cohort	Uphill steepness estimation Older adults judged 4 hills at two locations (hill set one were 2,3,4, and 25deg, hill set two were 3,5,20, and 29deg) Student aged group judged a single hill from a possible eight hills (2,4,5,6,10,21,31,33, and 34deg) Haptic, verbal, and visual matching measures
Age	Bian and Andersen Experiment 1	2012	Between group, cross sectional observational study	Older adults group n=9, 4(f); mean age 70.2 (4.9) years Young group n=8, 4(f); mean age 22.8 (2.4) years	Older vs younger adults	Distance estimation at 4,6,8,10, and 12m Verbal Action based measure (blind rope pulling)
Age	Bian and Andersen Experiment 2	2012	Between group, cross sectional observational study	Older adults group n=8, 4(f); mean age74.8 (5.1) years Younger group n=8, 4(f); mean age 21.9 (2.3)	Older vs younger adults	Distance estimation at 4,6,8,10, and 12m Verbal
Age	Bian and Andersen Experiment 3	2012	Between group, cross sectional observational study	Older adults group n=12, 6(f); mean age 76.2 (6.2) years Younger group n=12, 6(f); mean age 23.2 (2.8) years	Older vs younger adults	Distance estimation at 4,6,8,10, and 12m Verbal
Age	Bian and Andersen Experiment 4	2012	Between group, cross sectional observational study	Older adults group n=12, 6(f); mean age 71.8 (4.8) years Younger group n=12, 6(f); mean age 23.2 (2.8) years	Older vs younger adults	Distance estimation at 4,6,8,10, and 12m Verbal
Age	Costello et al. Experiment 2	2015	Between group, cross sectional observational study	Recruited n=114. Excluded 8 participants as outliers, and 2 from older group for failing mini-mental state examination (<27) After exclusion older adults group n=52, sex not reported; mean age 68.04 (range 60-80) years After exclusions younger group n=52, sex not reported; mean age 22.40 (range 18-33) years	Older vs younger adults. Health status (multi-dimensional health assessment questionnaire), fatigue, and gait speed were also measured for both groups	Distance estimation at 3.4, 7.9, 13.4, 20.4, and 25.3m Verbal
Age and Fitness	Dean et al.	2016	Exploratory cross-sectional study	n=106 were recruited from the surrounding community. 60(f), age range from 18-72 years	Association between Age and hill steepness estimation Also investigated associations with: experiential (hill) knowledge; fitness (body mass index [BMI])**; personality traits; and sex	Uphill steepness estimation of three hills (9, 22.5, and 4.5deg), always in the same order Verbal (all hills) Visual matching (9deg hill only) Haptic (9deg hill only)
Age	Eves et al. Experiment 1	2014	Cross-sectional observational study	n=269 participants were recruited at a train station. 129 (f), mean age 38.1 years (SE= 0.87), age range from 18-84 years	Associations between age and stair steepness estimation	Stair steepness estimations, from the base of the stairs and 15m away, the stairs were 23.4deg steep Verbal, visual matching, and haptic measures
Age	Norman et al.	2020	Between group, cross	Older group n=16, sex not reported; mean age 74.1 (SD=	Older vs younger adults	Distance estimation using an

			sectional observational study	6.1) years, age range from 65-83 years Younger group n=16, sex not reported; mean age 21.3 (SD= 3.2) years, age range from 19-22 years		equidistant cone task, where a cone was placed 6m away and participants were then asked to place 5 cones to create 5 distances that are the same to the first interval Visual matching
Age	Sugovic and Witt Experiment 1	2013	Between group, cross sectional observational study	Recruited n=29, excluded 5 from analysis as failed vision test (n=4 older group, n=1 younger group) Post exclusion older group n=12, 7(f); mean age 81.38 years Post exclusion younger group n=12, 7(f); mean age 20.27 years	Older vs younger adults The Physical Activity Questionnaire was used as a measure of functional capacity	Distance estimation at 4,6,8, and 10m Verbal
Body size	Bridgeman and Cooke Experiment 1	2015	Between group, experimental design	Recruited n=32. Excluded n=3 (all increased eye height group), due to guessing hills to be 90deg Increased eye height group (standing on a 37cm box): n= 17, age and gender not reported Control group (standing on ground): n=15, age and gender not reported	Eye height manipulated - participants stood on a 37cm box at the base of a hill Mean eye height of this group is 187.9cm (SD=9.63cm) compared to 140-160cm for the control group	Uphill steepness estimation (12deg) at 4 distances on the hill (2,4,8, and 16m) Verbal estimates
Body size	Bridgeman and Cooke Experiment 2	2015	Between group, experimental design. Replication of above with alternating allocation	Increased eye height group (standing on a 37cm box): n= 16, age and gender not reported Control group (standing on ground): n=17, age and gender not reported	Eye height manipulated - participants stood on a 37cm box at the base of a hill Mean eye height of this group is 187.9cm (SD=9.63cm) compared to 140-160cm for the control group	Uphill steepness estimation (12deg) at 4 distances on the hill (2,4,8, and 16m) Verbal estimates
Body size	Collier Chapter 5, Experiment1*	2017	Within group, experimental design	n= 36, 23(f), mean age 21.8 years	Hand size manipulated by wearing padded gloves. Each participant wore one padded, and one unpadded glove	Aperture width (4-14cm) estimated using each hand. Visual matching task
Body size	Collier Chapter 5, Experiment 2*	2017	Within group, experimental design. Replication of above with addition of a cover story	n= 36, 23(f), mean age 25.9 years	Hand size manipulated by wearing padded gloves. Each participant wore one padded, and one unpadded glove	Aperture width (4-14cm) estimated using each hand. Visual matching task
Body size	Jun et al.	2015	Between group, experimental design	Recruited n=42, n=1 excluded to having extremely large feet (>2SD from mean), n=3 due being outliers, and n=1 due to missing data n=37 analysed between both groups, no group sizes or demographics reported	Foot size was experimentally manipulated in virtual reality (VR). Feet rendered to be 51.9cm and 12.97cm in the large and small foot groups respectively	Distance estimations of a gap (0.5,0.9,1.05, and 1.7m) Verbal estimates
Body size	van der Hoort et al. Experiment 9	2011	Within group, experimental design. Replication of above with different measure of spatial perception	n=28, 16(f), mean age 26.4 (SE=1.2) years. All healthy adults naïve to the experimental hypothesis	Augmented reality (AR) "body swap" illusion, inducing embodiment of a large or small body	Distance estimation (8m) Action based measure (blind walking)
Body size	van der Hoort et al. Experiment 10	2011	Within group, experimental design	n=25, 9(f), mean age 28.4 (SE= 1.5) years. All healthy adults naïve to the experimental hypothesis	AR "body swap" illusion, inducing embodiment of a large or small body	Distance estimation (4,8, 16m) Verbal estimations
Embodiment	Phillips et al.*	2010	Between group, experimental design	Recruited n=12 for this study. 10 for avatar group and 2 additional male participants to add to a previous no avatar dataset Avatar group n=10, 1(f), aged between 18-33 years. No avatar group n=10, no demographics reported	Embodiment of an avatar vs. no avatar in a VR environment	Distance estimations; random distance between 2.4-6.1m Action based (blind walking) performance compared between VR and real world
Embodiment	Ries et al.*	2008	Between group,	Avatar group n=6 and no avatar group n=5. No	Embodiment of an avatar vs no avatar in a	Distance estimations (length not

			experimental design	demographics reported	VR environment	reported) Action based (blind walking) performance compared between VR and real world
Embodiment	Scandola *et al.* Experiment 2	2019	Within group, experimental design	n=20, 2(f), mean age 44.2 (SD=12.63). All had spinal cord injuries (SCI) and were manual wheelchair users (>6 months)	People with spinal cord injuries and use a wheelchair. Compared embodiment of own wheelchair vs other wheelchair	Distance estimation (2,3,4m), verbal estimation; and uphill ramp steepness (4,8, 16, 24, 32deg); visual angles. Both in VR
External load	Bhalla and Proffitt Experiment 1	1999	Between group, experimental design	Backpack group n=40, 20 (f), age not reported Control group n=90, 45 (f) The control group was previously collected as part of a normative dataset (Proffitt et al., 1995)	Experimental group wore a heavy backpack (~20% of body weight) Control group unencumbered	Uphill steepness measured at one of two hills (either 5deg or 31deg) Verbal, visual, and haptic measures
External load	Corlett *et al.*	1990	Between and within group, experimental design	Low resistance group n=8 High resistance group n=8 Overall mean age =22 years	An external load (high or low resistance) was applied through an elastic band attached to the participant's belt Each group (high or low) underwent 8 conditions where the resistance was applied or not applied	Distance estimation (9m) Action based measure (blind walking)
External load	Hutchinson and Loomis Experiment 1	2006	Between group, experimental design	Backpack group n=12, 6(f), age not reported Control group n=12, 6(f), age not reported	Experimental group wore a heavy backpack (1/5-1/6 of body weight) Control group unencumbered	Distance estimation (3,5,7,9,11,12, and 15m) Verbal estimation Action based measure (blind walking)
External load	Hutchinson and Loomis Experiment 2	2006	Within group, experimental design	n=12, 8(f), age not reported	Experimental condition wore a backpack (1/5-1/6 of body weight) Control condition was unencumbered	Distance estimation (3,8,11, and 15m) Verbal report
External load	Keric and Sebanz Experiment 3	2021	Within group, experimental design	Recruited n=20, 11(f), mean age 26.15 years Excluded n=3, 2x due to guessing purpose of the study, 1x due to making attempts to measure distance	Compared wearing a backpack that was 20% of body weight to an empty backpack Cover story used	Distance estimation (every meter from 1-14m) Verbal estimations
External load	Keric and Sebanz Experiment 4	2021	Within group, experimental design	Recruited n=20, 6(f), mean age 24.18 years Excluded 1x participant as they gave completely random estimates to finish quickly	Compared wearing a backpack that was 20% of body weight to an empty backpack No cover story used	Distance estimation (every meter from 1-14m) Verbal estimations
External load	Lessard *et al.*	2009	Within group, experimental design	Recruited n=18. Excluded n=2 due to not following instructions Post exclusion n=16, 8(f), age not reported	Experimental condition was ankle weights (5% of body weight) Control condition unencumbered	Distance estimations of gaps (0.15m- 3.05m) Verbal estimation
External load	Proffitt *et al.* Experiment 1	2003	Between group, experimental design	Backpack group n=12, 7(f), age not reported Control group n=12, 7(f), age not reported	Experimental group wore a heavy backpack (1/5-1/6 of body weight) Control group unencumbered	Distance estimation (4,6,8,10,12, and 14m) Verbal estimation
External load	Shea and Masicampo	2014	2 (backpack) x 2 (affirmation) factorial design	Total n=70, 59(f), mean age 19.0 (SD= 1.07) n of each group and group demographics not reported	Experimental group wore a heavy (20% body weight up to 11kg) backpack Control group wore and empty backpack	Distance estimation (39.6m) Verbal estimation
External load	Vinson *et al.* Replication condition	2017	Within group, experimental design	n=12, sex and age not reported	Experimental condition consisted of wearing a backpack that was 20% of body weight, followed by the control condition of no backpack	Distance estimation (8, 10, 12, and 14m) Verbal estimation
External load	White Experiment 1a*	2013	Within group, experimental design	n=22, 11(f), aged between 18-21 years	Compared two external load conditions (backpack manipulation and ankle weights both 10% of body weight) to a control condition (no external load)	Distance estimation (10, 12, 14m) Reproduce a previously walked distance on a treadmill in VR environment
External load	White	2013	Within group,	n=26, 9(f), aged between 18-26 years	Compared walking with two ankle weights	Distance estimation (10,12, 14m)

	Experiment 1b*		experimental design			conditions (5% and 10% of body weight) to a control condition (no external load)	Reproduce a previously walked distance on a treadmill in VR environment
External load	White Experiment 2*	2013	Within group, experimental design	n=17, 9(f), aged between 18-23 years		Compared three conditions of increased metabolic effort (condition of interest is ankle weights) to a control condition (normal walking)	Distance estimation (10,12, 14m) Reproduce a previously walked distance on a treadmill in VR environment
External load	White Experiment 4*	2013	Within group, experimental design	n=12, 7(f), aged between 19-28 years		Compared three experimental conditions (condition of interest ankle weights) to two control conditions (eyes open and eyes closed)	Distance estimation (12m) Blind-walking measure
Fatigue	Asaf et al.*	2015	Between (athletes vs. non athletes), within group (stationary vs. running on treadmill) experimental design	Athlete group n=15 and non-athlete group n=15 No other demographics reported		Compared performance of athletes and non-athletes before and during a fatiguing activity (treadmill running for 2min)	Distance estimation (12,18,24, and 32m) Visual matching
Fatigue	Baati et al.	2020	Within (normal sleep, first half of night, and second half of night sleep deprivation) x within (pre- and post-fatiguing exercise) experimental design	Recruited: n=10, football players, 0 (f), 22.8 (1.3) years		First half of night sleep deprivation (slept 3am-7am), second half of night sleep deprivation (slept 11pm-3am), compared to control night (slept 11pm-7am) Exercise condition was pre and post a repeated sprint cycling session	Distance estimation (15, 25, and 35m) Verbal estimations
Fatigue	Baati et al.	2015	Within (sleep deprivation) x within (pre- and post-fatiguing exercise) experimental design	Recruited n=10, football players, 0 (f), 22.8 (1.3) years		Sleep deprivation condition (0hrs sleep) compared to control night sleep (m=7hrs) Exercise condition was pre and post a repeated sprint cycling session	Distance estimation (15, 25, and 35m) Verbal estimations
Fatigue	Bhalla and Proffitt Experiment 2	1999	Within group, repeated measures	Recruited habitual runners, n=40, 20(f); age not reported		Pre and post exhausting run, which was between 45-75min long. Self-reported fatigue assessed	Uphill slope estimation of two hills (5deg and 31deg). Each participant estimated steepness of both hills, the order was counterbalanced. Verbal, visual matching, and haptic measures
Fatigue	Hunt et al.	2017	Between group experimental design	Opportunistic sample at a beach recruited n=83 Excluded n=1 (did not provide consent), final sample n=82 f(45), mean age 53.63 (16.64) Randomly assigned to exercise (n=41) and control (n=41) groups. No group demographics reported		Exercise group underwent a 90sec session of exercise on a stepper machine immediately prior to distance estimation task. Control group did nothing Manipulation check measuring self-reported fatigue found the two groups were statistically different (p=0.002)	Distance estimation (92.5m) Verbal estimation
Fatigue	Jarraya et al.	2013	Between group (athletes, sedentary, control) x within group (pre- and post-exercise) design	Athletes were all members of a professional soccer team, n=18, 0(f), mean age 23.7 years Sedentary group no participation in physical activity or sport, n=18, 0(f), mean age 23.8 years Control group were "normally active", n=18, 0(f), mean age 22.7 years		All groups underwent an exercise test (2min warm up, then 10min of cycling at 30-50% maximal aerobic potential) Spatial perception measured before, after 1-2min, 5-6min, 9-10min of exercise, and after the exercise test	Distance estimation (5,7,9, and 11m) all done at each timepoint Verbal estimation
Fatigue	Proffitt et al. Experiment 5	1995	Within group, repeated measures	Recruited n=60 (30[f]) university students that were regular runners		Pre and post exhausting run. No criteria on length or distance, just that participants	Uphill slope estimation of two hills (5deg and 31deg). Each participant

				Excluded 4x due to hill steepness knowledge, and 3x students that weren't fatigued by the running task Final sample n=53. No post-exclusion demographics provided	needed to be fatigued at the end of the run	estimated steepness of both hills, the order was counterbalanced. Verbal, visual matching, and haptic measures
Fatigue	Taylor-Covill and Eves Experiment 2	2013	Between group experimental design	Fatigued group = n=20, 9(f), mean age 19.83 (SD=1.58) years Control group n=20, 9(f), mean age 20.41 (SD-1.82)	Fatigued group participated in a maximal fitness test prior to spatial perception measure and control group did maximal fitness test after spatial perception measure	Slope estimation of a staircase (14.2deg) was projected on a wall Verbal, visual, and haptic measures
Glucose	Cole and Balcetis Experiment 1	2013	Between group, experimental design	Recruited n=54 undergraduates, 40(f) Scarce bioenergetics (low blood glucose) group n=27 Ample bioenergetics (high blood glucose) group n=27 n=1 participant excluded as outlier. Did not report which group this was from	"Bioenergetic" resources. Participants were given either a sugar drink or a placebo sugar drink. Blood glucose levels were measured after this manipulation	Distance estimation (length not reported) Action based measure (bean bag toss)
Glucose	Cole and Balcetis Experiment 2	2013	Between group, experimental design	Recruited n=72 undergraduates, 42(f), excluded n=2 from ample group as they guessed that the drink was a placebo Scarce psychoenergetics n=35 Ample psychoenergetics n=35	"Psychoenergetic" resources. Participants were all given non-caffeinated tea. The ample psychoenergetics group were told that the tea was a natural stimulant and the scarce psychoenergetics group were told that it was a sedative	Distance estimation (length not reported) Visual matching task
Glucose	Durgin et al.	2012	Between group, experimental design	Recruited n=39 undergraduates, 20(f) Did not report the n of each group	Participants were either given a sugar drink or a placebo drink	Uphill steepness (8.6deg) Verbal and haptic measures
Glucose	Schnall et al. Experiment 1	2010	Between group, experimental design	43 participants recruited, 7 excluded from analysis (2x didn't follow instructions, 5x outliers) Did not report the n of each group	Participants were either given a sugar drink or a placebo drink. Manipulation check found that participants were no better than chance of guessing their group	Uphill steepness (29deg) Verbal, visual, and haptic measures
Glucose	Zadra Experiment 1a*	2013	Between group, experimental design	n=43, healthy participants, fasted for 4hrs prior to experiment. Half were given a glucose drink, half a placebo drink. Sample size of each after exclusions (n=6) not reported	Blood glucose levels were measured at baseline, after glucose/placebo drink and Stroop test, and after spatial perception task	Distance estimation (6-10m in 0.5m increments) Visual matching
Glucose	Zadra Experiment 3a*	2013	Between group, experimental design	Recruited n=55, 30(f), mean age 18.92 (SD= 0.93) years. Divided into two groups: standard Jell-O (glucose) and placebo Jell-O (no glucose) Excluded 16 participants (1 due to procedural error, 1 due to drinking an extra drink during study, 1 for failure to follow instructions, 7 for previously participating in a slant study, and 6 as they ingested >10g of Jell-O)	Oral glucose levels. Participants were instructed to chew, but not swallow, either standard Jell-O or placebo (no glucose) Jell-O Participants self-rated fitness was also measured	Uphill steepness (5.6deg) Verbal and visual measures
Glucose	Zadra Experiment 3b*	2013	Between group, experimental design	Recruited n=71, 42(f), mean age 18.68 (SD=1.47) years. Dived into two groups: standard Jell-O (glucose) and placebo Jell-O (no glucose) Excluded 13 participants (2 due to extreme dislike of Jell-O, 3 for use of landmarks for distance estimation, 1 due to illness, 1 due to previously knowledge of study hypothesis, and 5 due to ingesting >10g of Jell-O)	Oral glucose levels. Participants were instructed to chew, but not swallow, either standard Jell-O or placebo (no glucose) Jell-O Participants self-rated fitness was also measured	Distance estimation (6,8,10,12m) Visual matching (line bisection task)
Glucose Weight/ Fitness	Schnall et al. Experiment 2	2010	Between group, experimental design (glucose manipulation) Association study (self-reported measures)	56 participants recruited, 10 excluded from analysis Did not report the n of each group	Glucose levels were measured before and after a glucose drink manipulation (sugar or placebo drink) Also used self-reported measures (bioenergetics test battery; self-reported sleep, exercise, nutrition, fatigue, mood, and stress)	Uphill steepness (5.6deg) Verbal, visual, and haptic measures

Glucose and external load	Shaffer et al.	2013	2 (backpack) x 2 (glucose) factorial design	4 groups (total n=120, 61 [f], no group demographics reported) placebo glucose, no backpack n=30 placebo glucose, backpack n=30 glucose, no backpack n= 30 glucose, backpack n= 30	Glucose levels were not measured in this sample. However, they did report the results of a pilot study that found that the glucose blood levels were statistically different in a previous comparison of the glucose and placebo groups	Uphill steepness (16deg) Verbal and haptic measures
Glucose	Zadra et al.	2016	Within (pre- and post-exercise) by within (glucose manipulation) experimental design	Recruited n=8 participants, 3(f), mean age 26.38 (SD=5.53) years	Compared spatial perception measured pre- and post-exercise on two occasions, once having ingested a glucose supplement and once ingesting a placebo supplement (no glucose)	Distance estimation Action based (blind walking)
Interoception	Mouatt et al.*	2021	Association design	Recruited and analysed 20 participants, 10(f), mean age 30.2 (SD= 11.2)	Associations between interoceptive accuracy and uphill steepness estimation. Used a heart rate accuracy task.	Uphill steepness (5-15deg) in virtual reality of 15 random hills Verbal estimates
Interoception	Tenhundfeld and Witt	2015	Between group, observational design	No details reported	Interoceptive accuracy - as measured by a HR accuracy task Split into high (2/3 of participants) and low (1/3 of participants) groups	Distance on hill
Pain	Tabor et al.	2016	Between group comparison, cross sectional	Recruited 72 participants (36 each group). Excluded n=10 (5 each group) due to variable distance units used by participants Chronic pain group n= 31, 27(f); mean age 43.3 (SD= 11.1) years Healthy control group n=31, 28(f); mean age 36.7 (SD= 13.7) years	Chronic pain group were all diagnosed with a chronic pain condition and recruited from a pain management centre. The average duration of pain was 12 (SD=9.7) years and diagnosis were: back pain (50%), CRPS (9%), multi-site (41%)	Distance estimation at 4,5,7,9, and 13m Verbal estimates
Pain	Witt et al.	2009	Between group comparison, cross sectional	Recruited n=20 participants (10 each group). Excluded 1x due to no pain during walking, and 2x due to being outliers Chronic pain group n=8, mean age 40.63 years Control group n=8, mean age 38.79 years	Chronic pain group consisted of people with MSK and neuropathic pain, average duration 9.02 (SD=8.00) years. All self-reported pain with walking	Distance estimation at 4,5,7, and 9m Verbal estimates
Pain	Alaiti et al.	2019	Between group comparison, cross sectional	Painful shoulder group n=84, 51(f); mean age 65(SD= 9.4) years Control group n=51, 35(f) ; mean age 57 (SD= 11.5) years	Chronic shoulder pain (right-handed and had pain in dominant shoulder for at least previous 3months)	Distance estimation to a point 45-100cm from body. Visual matching (ratio judgement)
Weight/ fitness	Bhalla and Proffitt Experiment 3	1999	Observational, between group design	total n=74 university students recruited, 35(f) age not reported. Recruited 24 varsity athletes, 8(f) and 50 unfit people, 27(f) to ensure wide range of fitness and weight outcomes 7 participants excluded due to missing data (fitness assessments)	Fitness level via a cycling test. Used HR and estimated VO2 max to create a fitness score from -2-4 BMI scores	Uphill steepness estimation of 4 hills (4, 5, 21 and 31deg) Verbal, visual and haptic measures
Weight/ fitness	Cole et al.	2013	Observational, between group design	Total n=78 participants recruited, 44 (f), age not reported. Split into low and high physiological potential groups Did not report the n of each group	Fitness measured using waist-to-hip ratio (deviation from "gender-specific ideals". Participants with high (1SD above mean) and low (1SD below the mean) hip-to-waist ratios were categorized into the low and high physiological potential groups respectively	Distance estimation (4.87m) Visual matching (on paper diagram)
Weight/ fitness	Krpan and Schnall Experiment 1	2014	Observational, between group design	Recruited n=54, 25(f), mean age 28.15 (SD=8.35). Excluded n=2 due to failure to follow instructions Split into low and high physical condition groups Did not report the n of each group	Self-rated fitness levels. Participants rated their current physical condition from 1 (very unwell) to 5 (excellent). Scores >1SD and <1SD from mean were the high and low fitness groups respectively	Uphill steepness estimation (39deg) Verbal, visual and haptic measures

Weight/ fitness	Krpan and Schnall Experiment 2	2014	Observational, between group design	Recruited n=58, 23(f), mean age 32.21 (SD=12.18). Excluded n=1 due to prior hill steepness knowledge Split into low and high physical condition groups Did not report the n of each group	Self-rated fitness levels. Participants rated their current physical condition from 1 (very unwell) to 5 (excellent). Scores >1SD and <1SD from mean were the high and low fitness groups respectively	Uphill steepness estimation (39deg) Verbal, visual and haptic measures
Weight/ fitness	Krpan and Schnall Experiment 3	2014	Observational, between group design	Recruited n=150, 52(f), mean age 27.6 (SD=6.92) years. Excluded n=5 for not completing all tasks, n=3 for not complying with instructions, n=1 for prior hill steepness knowledge Split into low and high physical condition groups Did not report the n of each group	Self-rated fitness levels. Participants rated their current physical condition from 1 (very poor) to 6 (excellent). Scores >1SD and <1SD from mean were the high and low fitness groups respectively	Uphill steepness estimation (39deg) Verbal, visual and haptic measures
Weight/ fitness	Shaffer *et al.*	2019	Association design	Recruited n=122 (75f), with a median age of 18. They were separated into 4 experimental groups; however, they also performed a secondary analysis of the correlation between self-rated fitness and spatial perception measures Excluded n=2 for steepness estimates over 90deg, and n=6 for guessing study hypothesis	Self-rated fitness levels Participants rated their physical fitness levels from 1 (not very physically fit) to 7 (very physically fit)	Uphill steepness estimation (6.1deg) Verbal Distance estimation (cone on hill 10m) Note - only done by 57 participants Verbal estimation
Weight/ fitness	Sugovic Experiment 2*	2014	Observational design, association study	Recruited n=73, 40(f), mean age 32.4 (SD=11.8) years. Excluded n=4 due to being morbidly obese, n=2 due to pregnancy, n=2 due to "muscular body type", and n=1 for being an outlier	Actual body size (BMI) and perceived body size (from modified Eating Disorders Questionnaire and figural scale)	Uphill steepness estimation (7deg) Verbal (not reported as 41% of participants had difficulty) Visual
Weight/ fitness	Sugovic *et al.*	2016	Observational, between group design	Recruited n=66, 30(f), mean age 24.4 (SD=6.53) years. Excluded n=2 due to not understanding task, n=2 due to experimenter error, n=1 as an outlier, n=1 as one estimate was much further than others, and n=4 for morbid obesity	Actual body size (BMI) grouped into normal (BMI 18.5-25), overweight (BMI 25-30), and obese (BMI 30-35) Perceived weight an evaluative measure of body size (too low - too high) and pictorial representation	Distance estimation (10,15,20, and 25m) Verbal estimation
Weight/ fitness	Taylor	2011	Observational, between group design	Traceur (athlete) group n=27, 0(f), mean age 19.9 years. Novices (control) group n=27, 0(f), mean age 19.5 years Second control group n=18, 0(f), mean age 20.1 years	Traceur group had a mean 16.77 months (SD=15.4) experience in parkour Controls were matched on age, sex, and height	Distance estimation of wall heights (1.94, 2.29, and 3.45m) Visual matching
Weight/ fitness	Taylor-Covill and Eves Experiment 1	2016	Observational, between group design	Cross-sectional study recruited n=187. Excluded n=7 due to belief that all stairs are 45deg, and n=9 for failing angle knowledge tests Post exclusion mean age 41.8 (SD=12.6), overweight group n=82, 41(f) and healthy group n=89, 42(f)	Experimenters used weight coding (BMI silhouettes) to separate into clearly healthy weight and clearly overweight groups	Staircase steepness (23.4deg) Verbal, visual and haptic measures
Weight/ fitness	Taylor-Covill and Eves Experiment 2	2016	Longitudinal cross-sectional study. Baseline assessment then a follow-up at 407 days (SD 137.45).	Recruited n=52 at baseline, 41(f), mean age 37.4 (SD=12.7) years Follow-up n=35. Lost to follow up due to health problems (n=4), pregnancy (n=2), relocation (n=2), and loss of contact (n=9) Report that there were no statistical changes in demographics between the baseline and follow-up participants	Body composition measured using a Hologic "Discovery" dual-energy X-ray aborptiometry body scanner	4 Staircase steepness estimations (between 20-33deg) Verbal, visual and haptic measures

2   Note: * refers to experiments that have not been peer reviewed, **refers to experiments that contain results that are relevant to more than one group, All reported SS is based on
3   that which was analysed unless otherwise stated.

Table 2: Risk of bias for included studies

For each category, green represents low risk of bias, yellow represents high risk of bias, red represents unsure risk of bias, and N.A., represents Not Applicable.

Study	Selection method	Sample size	Group Randomised	Task Randomised	Confounding variables	Manipulation achieved	Reliability	Participant blinding	Assessor blinding	Analysis blinding	Hypothesis a priori	Analysis a priori	Outcomes reported	Missing data <15%	Other
	Selection		Methodological					Blinding			Reporting				Other
Alaiti et al.	□	□	N.A.	□	□	□	□	□	□	□	□	□	□	□	
Asaf et al.	□	□	□	□	□	□	□	□	□	□	□	□	□	□	
Baati et al.	□	□	□	□	□	□	□	□	□	□	□	□	□	□	
Baati et al.	□	□	□	□	□	□	□	□	□	□	□	□	□	□	
Bhalla and Proffitt Experiment 1	□	□	□	□	□	□	□	□	□	□	□	□	□	□	
Bhalla and Proffitt Experiment 2	□	□	N.A.	□	□	□	□	□	□	□	□	□	□	□	
Bhalla and Proffitt Experiment 3	□	□	N.A.	□	□	□	□	□	□	□	□	□	□	□	
Bhalla and Proffitt Experiment 4	□	□	N.A.	□	□	□	□	□	□	□	□	□	□	□	
Bian and Andersen Experiment 1	□	□	N.A.	□	□	□	□	□	□	□	□	□	□	□	
Bian and Andersen Experiment 2	□	□	N.A.	□	□	□	□	□	□	□	□	□	□	□	
Bian and Andersen Experiment 3	□	□	N.A.	□	□	□	□	□	□	□	□	□	□	□	
Bian and Andersen Experiment 4	□	□	N.A.	□	□	□	□	□	□	□	□	□	□	□	
Bridgeman and Cooke Experiment 1	□	□	□	□	□	□	□	□	□	□	□	□	□	□	1
Bridgeman and Cooke Experiment 2	□	□	□	□	□	□	□	□	□	□	□	□	□	□	1
Cole and Balcetis 2013 Experiment 1	□	□	□	N.A.	□	□	□	□	□	□	□	□	□	□	2
Cole and Balcetis 2013 Experiment 2	□	□	□	N.A.	□	□	□	□	□	□	□	□	□	□	
Cole et al.	□	□	□	N.A.	□	□	□	□	□	□	□	□	□	□	
Collier Chapter 5, Experiment 1	□	□	□	□	□	□	□	□	□	□	□	□	□	□	
Collier Chapter 5, Experiment 2	□	□	□	□	□	□	□	□	□	□	□	□	□	□	
Corlett et al.	□	□	□	□	□	□	□	□	□	□	□	□	□	□	3
Costello et al. Experiment 2	□	□	N.A.	□	□	□	□	□	□	□	□	□	□	□	
Dean et al.	□	□	N.A.	□	□	□	□	□	□	□	□	□	□	□	4
Durgin et al.	□	□	□	□	□	□	□	□	□	□	□	□	□	□	4

Study	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Eves et al. Experiment 1			N.A.												
Hunt et al.				N.A.											
Hutchinson and Loomis Experiment 1															
Hutchinson and Loomis Experiment 2															
Jarraya et al.			N.A.												
Jun et al.															
Keric and Sebanz Experiment 3															
Keric and Sebanz Experiment 4															
Krpan and Schnall Experiment 1			N.A.												
Krpan and Schnall Experiment 2			N.A.												
Krpan and Schnall Experiment 3			N.A.												
Lessard et al.															
Mouatt et al.			N.A.												
Norman et al.			N.A.	N.A.											
Phillips et al.															
Proffitt et al. Experiment 1															
Proffitt et al. Experiment 5			N.A.												
Ries et al.															
Scandola et al. Experiment 2			N.A.												5
Schnall et al. Experiment 1															
Schnall et al. Experiment 2															
Shaffer et al. 2013															4
Shaffer et al. 2019															
Shea and Masicampo				N.A.											
Sugovic Experiment 2			N.A.												
Sugovic and Witt Experiment 1			N.A.												
Sugovic et al.			N.A.												
Tabor et al.			N.A.												
Taylor et al.			N.A.												

Study	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Score
Taylor-Covill and Eves 2013 Experiment 2				N.A.											
Taylor-Covill and Eves 2016 Experiment 1			N.A.												
Taylor-Covill and Eves 2016 Experiment 2			N.A.												
Tenhundfeld and Witt			N.A.												6
van der Hoort et al. Experiment 10				N.A.											
van der Hoort et al. Experiment 9															
Vinson et al.															
White Experiment 1a															7
White Experiment 1b															7
White Experiment 2															7
White Experiment 4															7
Witt et al.			N.A.												
Zandra Experiment 1a															
Zandra Experiment 3a				N.A.											
Zandra Experiment 3b															
Zandra et al.															

Details for Other Risk of Bias: 1) Risk of response bias as participants were repeatedly asked to estimate the same slope; 2) Likely poor reliability of spatial perception measured with action-based beanbag toss; 3) Manipulation of bodily state through experimenters adding external load via elastic bands, unlikely to be consistent within/between participants; 4) Scored low on reliability, however only one of the three measures of hill steepness has been assessed for reliability; 5) 8.9% of slope data contained estimations >90degs; 6) Primary author reported in correspondence that the group attempted to replicate the effect and failed (and these data have not been published –publication bias); 7) Potential bias in measurement of spatial outcome by assessor (unblinded assessor used a stopwatch to calculate distance walked).