

High-human acuity of speed asymmetry during walking

Pablo A. Iturralde Marcela Gonzalez-Rubio

Gelsy Torres-Oviedo

October 28, 2020

1 **Abstract**

2 Despite its central role in the proper functioning of the motor system, sen-
3 sation has been less studied than motor output in sensorimotor adaptation
4 paradigms. This deficit is probably due to the difficulty of measuring sensa-
5 tion: while motor output has easily observable consequences, sensation is by
6 definition an internal variable of the motor system. In this study we asked
7 how well can subjects estimate relevant environmental changes inducing mo-
8 tor adaptation. We addressed this question in the context of walking on a
9 split-belt treadmill, which allows subjects to experience distinct belt speeds
10 for each leg. We used a two-alternative forced-choice perceptual task (2AFC)
11 in which subjects report which belt they thought to be moving slower. We
12 characterized baseline accuracy in this task for healthy human subjects, and

13 found 75% accuracy for 75 mm/s speed differences. Additionally, we used a
14 drift-diffusion model of the task that could account for both accuracy and
15 reaction times. We conclude that 2AFC tasks can be used to probe sub-
16 jects' estimates of the environment and that this approach opens an avenue
17 for investigating perceptual deficits and its relation to motor impairments in
18 clinical populations.

19 **2 Introduction**

20 Despite its central role in the proper functioning of the motor system, sen-
21 sation has been less studied than motor output in sensorimotor adaptation
22 paradigms. This deficit is probably due to the difficulty of measuring sensa-
23 tion: while motor output has easily observable consequences, sensation is by
24 definition an internal variable of the motor system. For example, there exists
25 abundant literature characterizing the adaptation of motor behavior evoked
26 by split-belt walking (i.e., legs moving at different speeds) under a variety of
27 conditions (e.g. Dietz et al., 1994; Torres-Oviedo and Bastian, 2010; Ogawa
28 et al., 2014; Mawase et al., 2014), and for different populations (e.g. Reisman
29 et al., 2007; Finley et al., 2015; Sombric et al., 2017), but the contribution of
30 sensory information to this task is less understood. Psychophysical studies,
31 in which subjects report what they perceive in response to a given physical
32 stimuli, have been used to investigate the sensed speed differences that drive
33 motor adaptation during split-belt treadmill walking (e.g. Jensen et al., 1998;
34 Lauzière et al., 2014; Hoogkamer et al., 2015; Wutzke et al., 2015; Vazquez
35 et al., 2015; Statton et al., 2018; Leech et al., 2018). However, the method-

36 ological approaches followed by these studies might be limited for the purpose
37 of quantifying sensation.

38 Current literature on perception of speed differences fails to consider fac-
39 tors influencing people’s responses such. For example, people can report
40 difference perception for the exact same sensory stimulus when probed mul-
41 tiple times (i.e., the probabilistic nature of perceptual responses). Similarly,
42 the time it takes for people to generate a response varies depending on how
43 confident the subject is (i.e. subject-specific confidence in their responses).
44 These factors are important for inferring human sensation from percep-
45 tual tests (Ehrenstein and Ehrenstein, 1999). During split-belt walking the
46 external stimulus which needs to be sensed, which induces adaptation by
47 disrupting the gait pattern, is the speed difference between the legs.

48 Previous reports have assessed people’s perception of belt speed differ-
49 ences through variants of yes/no tasks, in which individuals report when they
50 perceive the belts to move at different speeds (i.e. yes, there is a speed differ-
51 ence) or the same speed (i.e. no, there is not a speed difference). Regardless
52 of the sensory stimuli presentation modality, perceptual responses to yes/no
53 tasks depend not only on subjects’ sensation, but also on subject-specific
54 decision criteria to convert sensory information into a response (Ehrenstein
55 and Ehrenstein, 1999). For example, some subjects are more likely to fa-
56 vor a response (e.g. ”yes”) when in doubt. Resulting in different responses
57 to identical sensory information for different subjects. Thus, the impact of
58 subject-specific decision criteria in these yes/no assessments suggests that
59 other perceptual methods, such as the two-alternative forced-choice task,
60 might be preferable to study sensation.

61 The two-alternative-choice task (2AFC) is preferable to yes/no tasks be-
62 cause it allows us to access what people feel despite of their confidence on
63 the stimuli that they are detecting. Notably, on a 2AFC task, subjects are
64 asked to judge which of two stimuli contains a certain signal or satisfies a
65 certain property. Importantly, subjects are forced to respond between the
66 two alternatives, regardless of their confidence on the decision. One might
67 consider that it is easier to indicate that something is happening (e.g., yes,
68 there is a difference), rather than what is happening (e.g., left side mov-
69 ing slower than right one). However, the sensation thresholds determined
70 through 2AFC tasks tend to be smaller in magnitude than those determined
71 through yes/no tasks (Green and Swets, 1966). Consistently, studies using
72 the two-alternative forced choice (2AFC) task that we propose have shown
73 that individuals can detect external stimuli without explicit awareness of said
74 stimuli (Goldstein, 2009; Ehrenstein and Ehrenstein, 1999). Thus, while the
75 declarative aspect of our perceptual task is undeniable, the 2AFC task will
76 give us a metric of at least partial sensory information available for sensori-
77 motor recalibration.

78 Forcing a choice is a method to infer subjects' sensations that would not
79 get reported with other perceptual tests. Critically, 2AFC eliminates subject-
80 specific differences in value between the alternative responses that exist for
81 the yes/no task, making it simpler to obtain a quantification of sensor acuity.
82 In other words, the alternatives in a yes/no task (i.e., "yes" or "no") are
83 not necessarily valued in the same way by every person. Thus, individuals
84 may set personal decision thresholds adjusting for the value given to each of
85 them. In contrast, in the 2AFC method the two alternatives are presented in

86 a symmetric way, such that there is no additional value in one choice over the
87 other. This makes 2AFC the preferred methodological approach to obtain a
88 measure of sensory acuity (Green and Swets, 1966).

89 While detecting accuracy is considered to be the main outcome measure
90 of perceptual tasks, reaction times are key in the process of accumulation of
91 sensory evidence before making a decision in a discrimination task (Pardo-
92 Vazquez et al., 2019; Henmon, 1911). Moreover, by using a 2AFC task we
93 have access to information on subjects' reaction times, which indicates the
94 available sensory information when analyzed through an appropriate model
95 of how choices are made. One such mechanistic model that can systematically
96 explore both accuracy and reaction times in a discrimination task is the drift-
97 diffusion model (e.g. Ratcliff, 1978; Bogacz et al., 2006; Gold and Shadlen,
98 2007). In the drift-diffusion model (DDM) for 2AFC an evidence variable
99 is used to represent the accumulation of information, and a choice is made
100 when one of two alternative choice barriers, corresponding to the two possible
101 responses, is reached. Thus, the DDM offers a principled way to link accuracy
102 and reaction times in the 2AFC as two expressions of the same mechanism
103 for gathering sensory evidence to make a choice.

104 Here we rigorously characterized the human ability to detect differences in
105 belt speeds on a split-belt treadmill. The focus of our study was to evaluate
106 sensory information available to subjects, rather than their confidence levels
107 or other decision criteria involved in eliciting responses. Consequently, we
108 used a 2AFC task for our perceptual assessment. We present quantifications
109 of accuracy, reaction times, and estimates of perceptual thresholds across
110 subjects and stimuli magnitude. Further, we used a drift-diffusion model

111 to gain insight into the processes underlying subjects choices and reaction
112 times. Our results may be used as normative data to compare to other pop-
113 ulations whose sensory acuity may differ or to assess changes in perception,
114 for example, as a consequence of sensorimotor recalibration or lesions to the
115 nervous system.

116 **3 Methods**

117 **3.1 Data Collection**

118 **Participants** $N = 9$ healthy subjects (24.6 ± 3.7 y.o., 6 female) com-
119 pleted the protocol. All of the subjects were right-footed (self-reported leg
120 used to kick a ball) and two of the subjects were left handed (self-reported).
121 The protocol was approved by the University of Pittsburgh’s Internal Review
122 Board (IRB) in accordance to the declaration of Helsinki.

123 **Testing protocol.** The overall protocol subjects experienced is depicted
124 in Figure 1A. Throughout the whole protocol, participants walked on a split-
125 belt treadmill with a mean speed between their legs of $1.05m/s$. The subjects
126 were initially familiarized with the perceptual task they were going to be per-
127 forming throughout the protocol by performing 6 repetitions of the task while
128 being provided with both visual (a live graphic of both belts’ speeds) and
129 verbal feedback from the experimenters. This ensured subjects understood
130 the task and mapping between their actions (key-presses) and the changes
131 in the speed of the belts. Following familiarization, subjects performed 2
132 to 4 blocks of data collection (as time permitted). Each block consisted of
133 interleaved tied-belt walking (i.e., both belts move at the same speed for 25
134 strides) and the perceptual tasks at regular intervals. The blocks had a single
135 presentation of the non-zero stimulus sizes, defined as an imposed belt-speed
136 difference ($\Delta v = v_R - v_L$) at the beginning of each perceptual trial, and two
137 presentations of the null trials in pseudo-random order (the same order for all
138 subjects). There was a total of 24 trials per block (see figure 1B), where the
139 stimulus sizes consisted on any of the following values: 0 (null), ± 10 , ± 25 ,

140 ± 50 , ± 75 , ± 100 , ± 125 , ± 150 , ± 200 , ± 250 , ± 300 , and ± 350 mm/s. The
141 treadmill was only stopped between blocks, but not within them, and sub-
142 jects were allowed to rest and move as desired during the breaks. Even blocks
143 were mirror images of the odd blocks, so if the $+250\text{mm}/s$ was presented first
144 in the odd blocks, then the $-250\text{mm}/s$ was presented first in the even blocks
145 and so on. This ensured balancing of positive and negative perturbations to
146 minimize experimentally-introduced biases in responses. We collected a total
147 of $M = 30$ blocks from the $N = 9$ subjects, with two subjects completing
148 just two blocks, two subjects completing 3, and the rest completing 4 blocks
149 each.

150 **Description of the perceptual task.** The perceptual task was de-
151 signed to assess subjects' perception of speed differences based on two meth-
152 ods: 2-alternative forced-choice followed by speed-matching (see Figure 1,
153 panel C). The speed-matching component was a variant from previous per-
154 ceptual tasks (Jensen et al., 1998; Vazquez et al., 2015) and is not analyzed
155 in this study. Results from this portion of the task will be used in a future
156 study comparing methods to track shifts in perception over motor adapta-
157 tion. Every perceptual trial started with subjects walking with both belts
158 moving at $1.05\text{m}/s$ followed by a sudden transition in belt-speeds to a speed
159 difference whose value was unknown to the subjects (i.e., stimulus size). Sub-
160 jects walked at this speed difference for a full stride cycle (i.e., time duration
161 between two foot landings of the same leg) after which they heard an audio
162 cue signaling the beginning of the response window (Figure 1C gray shaded
163 area). Upon this audio cue, subjects had to press one of two keys (left or
164 right) according to which belt they perceived to be moving slower. Response

165 time (i.e., reaction time) and accuracy in the first key-press was used for the
166 2-alternative forced choice analysis. Subjects were also asked had to repeat-
167 edly press either key until the two belts felt like they were moving at the same
168 speed (speed-matching component). At every key-press, the belt speed dif-
169 ference would change by either 6, 8 or 10 mm/s (equally probable, randomly
170 chosen) such that the belt speed difference was either reduced (e.g., the sub-
171 ject pressed the key corresponding to the belt that was moving slower) or
172 increased (e.g., the subject incorrectly pressed the key corresponding to the
173 belt that was moving faster). Changes in belt speed were split symmetrically
174 across the two belts such that the mean speed across the belts was constant
175 throughout the experiment (1.05 m/s). The response window lasted for 24
176 strides, after which they would hear a second, different, audio cue indicating
177 the end of the perceptual task. Subjects wore noise-cancelling headphones
178 and a drape that blocked vision of their feet throughout the protocol to re-
179 move any additional auditory or visual information influencing the responses.
180 The headphones were also used to provide the start/stop audio cues, as well
181 as a clicking sound at each key-press to avoid any impression that the system
182 may not be detecting their actions.

183 **3.2 Perceptual Trial Exclusion**

184 Trials where subjects did not respond and the first trial in each block were
185 excluded from analysis. Non-response trials consisted of 1.1% of all recorded
186 trials (8 out of 720), and occurred in small stimulus size trials only (i.e., 5/60
187 of ± 25 mm/s trials, 1/60 of ± 50 mm/s trials and 2/120 of ± 0 mm/s trials).
188 We also eliminated the first perceptual trial for every block because we re-

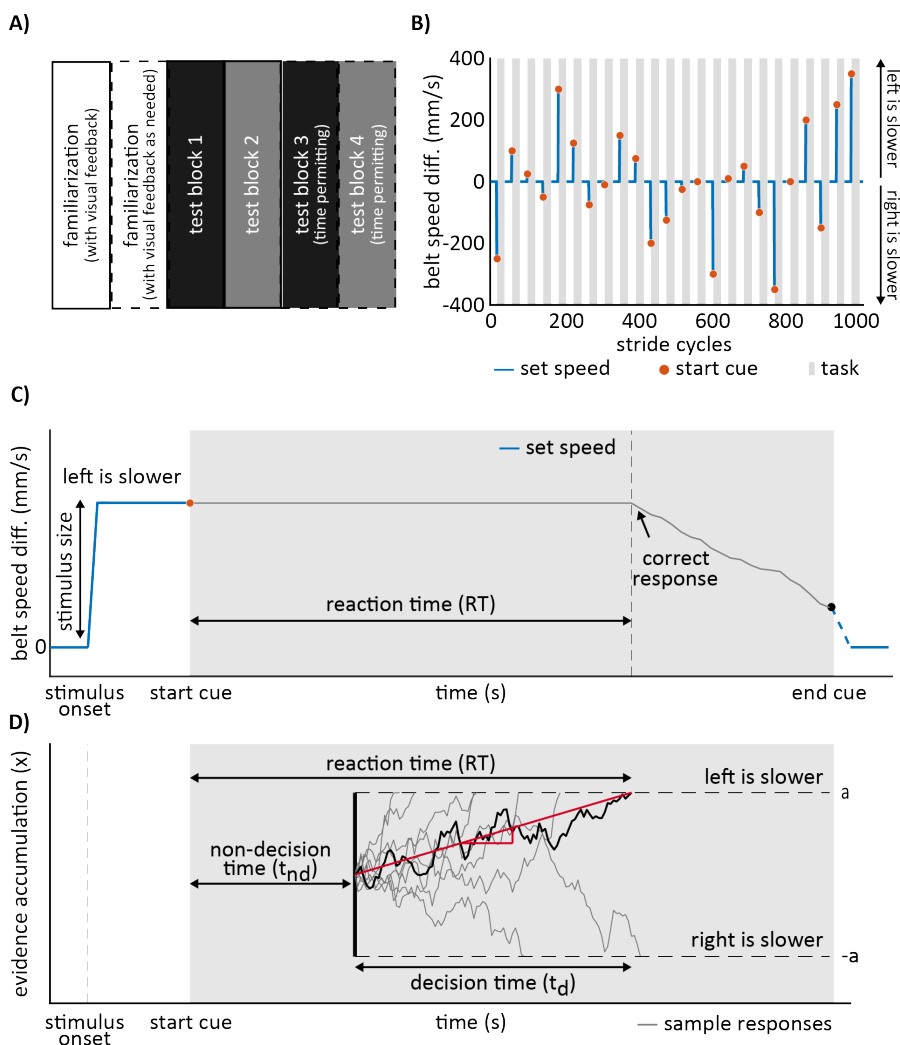


Figure 1: Protocol and methods for characterizing the perception of belt speed differences. **A)** Experimental protocol. Subjects completed one or two familiarization blocks followed by 2 to 4 testing blocks. **B)** speed difference profile for odd testing blocks. The profiles took the opposite values for the even blocks. Subjects performed a perceptual task in each of the shaded intervals. **C)** Perceptual task description. Tasks began and ended with audio cues. Upon hearing the cue, subjects were instructed to identify the belt moving slower, and to make as many key-presses as necessary until the belts felt as moving at equal speeds. **D)** Drift-diffusion model schematic. The drift-diffusion model for 2AFC tasks represents the temporal evolution of a decision variable as a random-walk (black and gray jagged lines). Decisions are made when one of the two decision barriers is reached (dashed lines). Reaction time is composed of a non-decision and decision interval.

189 alized that other factors (e.g., distraction) beyond actual perception had an
190 impact on the initial accuracy metric for every block. For example, subjects'
191 accuracy probed at 250 mm/s (which is a very salient speed difference) was
192 much smaller when presented at the beginning than within the block (64%
193 and 81% respectively).

194 **3.3 Perceptual Task Outcome Variables**

195 From each presentation of the perceptual task, we extracted two outcome
196 variables that were subsequently used for analysis, both from the two-alternative
197 forced-choice portion of the task: choice (first key pressed) and reaction time
198 (time until first keypress). Results from the speed-matching portion of the
199 task are not analyzed here. A 'left' choice was defined as a 'left is slower
200 than right' response. A 'right' choice is defined analogously. Choices were
201 converted to accuracy scores for some analyses. A response was considered
202 accurate if the subjects' choice indicated correctly the belt that moved slower.
203 Accuracy was not defined for null trials (i.e. 0 mm/s stimulus size), as there
204 is no correct choice. Reaction time was defined as the interval between the
205 starting audio cue and the first keypress. Different from accuracy, reaction
206 time is defined for all trials, including null ones.

207 **3.4 Quantification of Perception Through the 2AFC** 208 **Task**

209 **Perceptual thresholds (JNDs) and point of subjective equality (PSE).**

210 We used the 2AFC task to estimate two important quantities in the charac-

211 terization of perception: the just noticeable difference or differential threshold
212 (JND), and the point of subjective equality (PSE). The JND or threshold is
213 defined as the minimum magnitude difference between two stimuli that can
214 be reliably perceived by the subject as distinct (e.g. Green and Swets, 1966).
215 If we adopt a deterministic view of sensation, where a specific stimulus is
216 either always detected or never detected, this definition matches the notion
217 of a transition point between detection and non-detection. However, this all
218 or nothing assumption rarely conforms to observations in perceptual stud-
219 ies, (Ehrenstein and Ehrenstein, 1999), requiring a probabilistic definition
220 of thresholds. Thus we adopt the midpoint threshold definition (Goldstein,
221 2009), which is the stimulus value when probability of detection first exceeds
222 the midpoint between its lowest and its largest possible values. In a 2AFC
223 task such as the one used here, 75% accuracy is commonly used, given that
224 subjects can be 50% accurate from merely guessing at the task. The point
225 of subjective equality (PSE) is defined as the stimulus for which subjects
226 choose equally between two options (50/50 choice ratio). This is a metric
227 that quantifies potential biases in sensation (i.e., subjects are more likely to
228 choose left than right or vice versa) and has been used in studies characteriz-
229 ing changes in perception following split-belt walking (Vazquez et al., 2015;
230 Statton et al., 2018; Leech et al., 2018).

231 Both PSE and JND can be directly estimated by finding the stimuli
232 values for which subjects select between the two alternatives with 50/50 and
233 75/25 ratios. While this is feasible in principle, accuracy estimates for each
234 stimulus are very noisy, especially at the individual level were each stimulus
235 was presented at most 4 times. Consequently, we decided against the direct

236 approach and used responses for all stimuli presented to obtain a smooth
237 estimate of the relationship between stimuli and responses in the task (i.e.,
238 psychometric function).

239 **Psychometric curve fitting.** As stated in the previous section, esti-
240 mation of the PSE and JND was predicated on first characterizing the choice
241 vs. stimulus size curve in this task. We did this through a maximum like-
242 lihood fit to the binary responses for both individual and group-averaged
243 data. Specifically, we fit the binary left/right responses (not accuracy) data
244 from each individual using a parametric logistic regression approach. The
245 responses were modeled as coming from a binomial distribution with param-
246 eter p , where said parameter was taken to be a logistic function of a linear
247 combination of the factors of interest (summarized in the function μ), as
248 shown in Eq. 1.

$$p = \frac{1}{1 + e^{-\mu}} \quad (1)$$

249 Initially, we considered two main factors that explain subject choices at
250 each trial: a bias (intercept) term, and the stimulus size (ΔV_k). We then
251 considered the possibility that three exogenous factors may affect subject
252 choice and consequently our JND estimate: task learning, habituation, and
253 dominance. To control for these spurious effects we included three additional
254 factors to our regression: previous stimulus size (ΔV_{k-1}), a current stimulus
255 by block number interaction ($\Delta V_k \times b_j$, where $b_j \in [0, 1, 2, 3]$ indicates the
256 block number), and a term that depends solely on absolute stimulus size
257 ($|\Delta V_k|$). This results in a five-factor model as shown in Eq. 2.

$$\mu_k = \beta_0 + \beta_1 \Delta V_k + \beta_2 \Delta V_{k-1} + \beta_3 \Delta V_k \times b_j + \beta_4 |\Delta V_k| \quad (2)$$

258 The intercept (β_0) term quantifies a potential bias in responses. The sec-
259 ond term (β_1) quantifies sensitivity to stimulus size. The third term (β_2)
260 quantifies subjects' habituation to previously presented belt speed differ-
261 ences. That is, because of prior experience subjects responses may be chang-
262 ing. In its simplest form, this can be taken as the responses to each stimulus
263 being affected by the immediately preceding stimulus size. A negative value
264 represents that having two stimuli of the same sign makes the second one
265 more difficult to identify, while a positive value represents the opposite. The
266 fourth term (β_3) quantifies learning in the task, such that subjects would have
267 a better performance in perceptual tasks presented later in the experiment.
268 A significant value of β_3 indicates a change in the slope of the psychometric
269 functions for different blocks. A positive value would represent subjects get-
270 ting increasingly better at the task (sharper transition between left and right
271 choices) which is the expected effect, if any. Finally, the absolute stimulus
272 size term (β_4) would indicate a dominance effect. That is, it would indicate
273 a higher sensitivity to one specific belt moving faster.

274 The model was first fit to all responses (pooled across subjects), to select
275 the relevant factors among those considered. The model fitting was done
276 using Matlab's (The Mathworks, Inc., Natick, Massachusetts, United States)
277 *fitglm* function. A stepwise procedure was used to drop non-significant terms
278 from the model one at a time. The criterion for dropping terms was a p-value
279 larger than 0.05 for the likelihood ratio test of the model with and without
280 the corresponding term, under a χ^2 distribution with one degree of freedom.

281 This procedure is implemented by Matlab’s *stepwiseglm* using the deviance
282 criterion. Group regression results showed that only the bias and stimulus
283 size factors (β_0 and β_1) were significant. Consequently, we fitted individual
284 choice models considering those two factors only, as shown in Eq. 3.

$$\mu_k = \beta_0 + \beta_1 \Delta V_k \quad (3)$$

285 **Estimation of the JND.** Given the psychometric fits described above, it
286 can be shown that for an unbiased subject 75% accuracy happens at stimuli
287 values of $\approx \pm 1.1\beta_1^{-1}$. Consequently, we use the estimate $\text{JND} = 1.1\beta_1^{-1}$.
288 More generally, this JND quantification can be interpreted as the increase in
289 stimulus needed to go from 50/50 response proportion to a 75/25 proportion
290 (in favor of either response) if all other parameters and factors are held
291 constant.

292 **Estimation of the PSE.** For the same psychometric fits described
293 above, the belt speed difference for which subjects show equal proportion
294 of responses is $\text{PSE} = -\beta_0/\beta_1$. We note that for healthy subjects the PSE
295 in this task is expected to be close to 0 mm/s, such that if both belts are
296 moving at the same speed, half of the time subjects will choose left and the
297 other half right.

298 **3.5 2AFC Decision Making as a Drift-Diffusion Pro-** 299 **cess**

300 The drift-diffusion model (DDM) is a model of decision making with noisy
301 evidence. In this model, subjects are assumed to accumulate evidence in

302 time until sufficient evidence is gathered and a decision is made. Here we
303 consider the simplest form of DDMs for a two-alternative choice task. The
304 evidence gathered up until time t is represented as the continuous variable
305 $x(t)$. Whenever $x(t)$ goes above the barrier $a(t)$ we say that a left choice
306 has been made, and whenever it goes below $b(t)$ we say a right choice has
307 been made (see Figure 1D). Whenever the first barrier crossing happens, the
308 trial terminates. We assume starting point is unbiased (i.e. that the starting
309 point is equidistant to both decision thresholds, and thus subjects have no
310 preference for either response), and fixed decision barriers in time. Thus,
311 without loss of generality we take the starting point of the decision variable
312 to be $x = 0$ and the barriers taken to lie symmetrically so $a(t) = -b(t) = a$.
313 The simplest DDM can then be characterized by three additional parameters:
314 the noise level (σ), the drift-rate (r), and the non-decision time (t_{nd}). The
315 model separates the evolution of $x(t)$ into two stages. First there is a non-
316 decision stage, representing a delay (t_{nd}) in the beginning of the evidence
317 gathering (Eq. 4).

$$x(t) = 0, t \leq t_{nd} \quad (4)$$

318 In the following stage the evidence gathering process is modeled as a contin-
319 uous stochastic process with the following evolution (also known as a Wiener
320 or Brownian motion process with drift):

$$dx = r.dt + \sigma.dw, t > t_{nd} \quad (5)$$

321 Where r and σ are constants (for a given stimulus), dx refers to the change in
322 accumulated evidence for an infinitesimal time interval dt , and dw is a zero-

323 mean normal process such that $dw \sim N(0, dt)$ for that same time interval.
324 This equation can be interpreted as a process that accumulates evidence
325 linearly (in time) through the given drift rate r , but is affected by additive
326 noise also accumulated in time.

327 We note that despite there being four parameters for the model (a , σ ,
328 r , and t_{nd}) the model is scale-invariant, such that proportionally scaling the
329 values of a , σ , and r by the same amount results in the same predicted
330 behavioral outcomes. Hence, without loss of generality we define "a" to be
331 equal to 1 (Wagenmakers et al., 2007). Then the probability of the process
332 hitting one particular decision barrier (e.g., the probability of a left choice
333 being made, P_L , which corresponds to the positive or upper decision barrier;
334 Figure 1D), and the mean reaction time have closed-form expressions (Bogacz
335 et al., 2006; Wagenmakers et al., 2007) given by:

$$P_L = \frac{1}{1 + e^{\frac{2r}{\sigma^2}}} \quad (6)$$

336

$$mean\ t_d = \frac{1}{2r}(2P_L - 1) \quad (7)$$

337 By extension, the probability of hitting the other barrier is $P_R = 1 - P_L$. We
338 note that the choice probability in the task is fully determined by r and σ ,
339 and satisfies $\frac{2r}{\sigma^2} = \log\left(\frac{1-P_L}{P_L}\right)$.

340 Given the protocol design, where subjects experience the decision task
341 for different stimulus sizes, it is necessary to establish the dependency of the
342 model parameters (r , σ , t_{nd}) to the different experimental conditions (i.e.,
343 different stimulus sizes). The experimental condition kept the mean speed
344 as a fixed number ($v = \frac{v_R + v_L}{2}$). Therefore, it is possible to simplify the

345 analysis on the dependency of the model parameters and the stimulus size
346 such that the stimuli will be characterized jointly through their difference
347 $\Delta v = v_R - v_L$. Furthermore it is assumed that: 1) The non-decision time t_{nd}
348 is independent of Δv . 2) Noise or diffusion rate is symmetric; that is, that
349 a transposition of v_R and v_L leads to the same diffusion rate. The simplest
350 such relation is to assume that the diffusion rate σ depends on the stimuli
351 as $\sigma = \sigma_0 + k\Delta v^2$, where σ_0 is a baseline noise and k scales the dependence
352 on the stimuli. This relation can be interpreted as a second-order (Taylor)
353 approximation of a more general relation that is symmetrical on the stimuli
354 (i.e., that the diffusion rate is invariant to flipping the speeds of the two belts).
355 3) Finally, we assume that choice in the task must scale with stimulus size.
356 Because choice is completely determined by $\frac{2r}{\sigma^2}$, the simplest such relation is
357 $\frac{2r}{\sigma^2} = b + c\Delta v$, where b represents a bias term and c a scaling term. The drift
358 rate r is then implicitly related to the stimuli. Similar to before, this can be
359 interpreted as a first-order approximation of a more general dependency of
360 $\frac{2r}{\sigma^2}$ on Δv . We note that the model then has five scalar degrees of freedom
361 (b , c , t_{nd} , σ_0 , and k).

362 Using these definitions, the model was fit through a two-stage procedure:
363 First, we find the maximum likelihood fit to choices (left/right), which de-
364 termines the parameters b and c (Eq. 6). Second, we perform a least-squares
365 fit to mean reaction times, which results in estimates of t_{nd} , σ_0 and k (Eq.
366 7). We will consider the special case with absence of signal-dependent noise
367 ($k = 0$ and $\sigma = \sigma_0$ is a constant), where the drift in the DDM is the only
368 model parameter dependent on the stimulus size. Moreover, we analyzed the
369 case in which $k \neq 0$, where both the drift and the diffusion term in the DDM

370 are dependent on the stimulus size (i.e., there is signal-dependent noise). Pa-
371 rameters were fit to each individual separately, but the resulting models are
372 presented by averaging across all subjects for visualization purposes.

373 **3.6 Data and Code Availability**

374 Perceptual data and the code used for all analyses and creation of figures in
375 this work are available at [THIS HAS BEEN ANONYMIZED, WILL RE-
376 PLACE WITH PROPER URL AFTER MANUSCRIPT ACCEPTANCE]Kinematic
377 and kinetic data from subjects while performing the task is available upon
378 request.

379 4 Results

380 4.1 Characterization of Sensation Through the 2AFC 381 Task

382 To answer the question of what factors could influence subjects' perception
383 of the speed at which their legs moved, subjects performed the 2 alterna-
384 tive forced choice task described in the methods. We fitted a logistic re-
385 gression function to the averaged pooled responses (Figure 2A, black solid
386 line), considering the effect of several factors beyond stimulus size. How-
387 ever, only stimulus size ($\beta_1 = 0.012 \pm 0.001$, mean \pm standard error) sig-
388 nificantly influenced subjects' choices ($p = 2.45 \times 10^{-34}$). The intercept coef-
389 ficient ($\beta_0 = 0.192 \pm 0.098$, mean \pm standard error) almost reached signif-
390 icance ($p = 0.051$), indicating a potential group bias. However, this effect
391 was mainly driven by a single subject with a large bias, and thus, should be
392 interpreted carefully. This term is kept in the model for better fitting of the
393 data on an individual level. Lastly, the other factors that were studied did
394 not significantly affect subjects' responses, such as leg dominance ($p = 0.93$),
395 previous stimulus size ($p = 0.25$), or stimuli habituation (i.e., stimulus size
396 by block number interaction, $p = 0.17$).

397 **JND (group level)** The JND was estimated by smoothing available
398 data and finding the point at which the logistic model, fitted to the subjects'
399 pooled data, crossed the 75% accuracy value. The group-level threshold (see
400 methods) is 95 mm/s (equivalent to a 9.1% Weber fraction). The Weber
401 fraction for this sensory modality is given by the threshold as a fraction
402 of the mean belt-speed (i.e., $w = \Delta v / \bar{v}$). We note that the grouped data

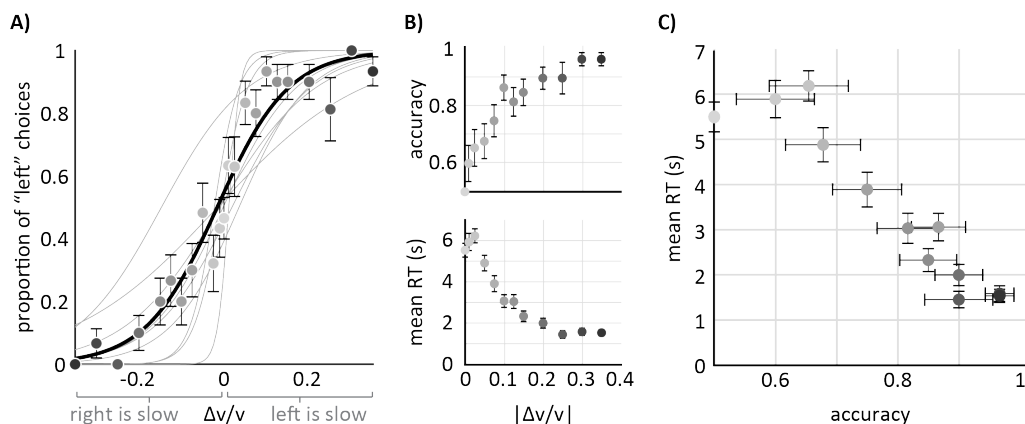


Figure 2: Choice, accuracy, and mean reaction times vs. stimulus size. **A)** choice as a function of stimulus size. Circles indicate group average responses across subjects (\pm standard error of the mean), thin gray lines represent logistic fits to individual data (see Methods), thick black line represents the average of the individual logistic fit curves (not the logistic fit to group averaged data). **B) Top:** accuracy as a function of absolute stimulus size. Both the data and model fits are the same as in the left panel, but averaged across positive and negative stimuli. Circles indicate experimental data (group average \pm standard error). **Bottom:** mean reaction time as a function of absolute stimulus size. Circles indicate experimental data (group average \pm standard error). **C)** mean reaction time vs. accuracy. Circles indicate experimental data (group average \pm standard error). Note that the color gradient in the circles among B and C depend on the absolute stimulus size.

403 estimated over the 75 mm/s stimulus, has a group accuracy of exactly 75%
404 (Figure 2B, top panel). We conclude that the perceptual threshold defined by
405 75% accuracy in the task is on average located approximately at 7% (Weber
406 fraction) across the population tested.

407 **Individual differences in accuracy, PSE, and JND** Individual sub-
408 jects displayed a large range of behaviors in the 2AFC, with accuracy varying
409 from 65% to almost 95% for individual subjects across all trials (see Figure
410 3A. Large reaction time differences were also observed, with mean values

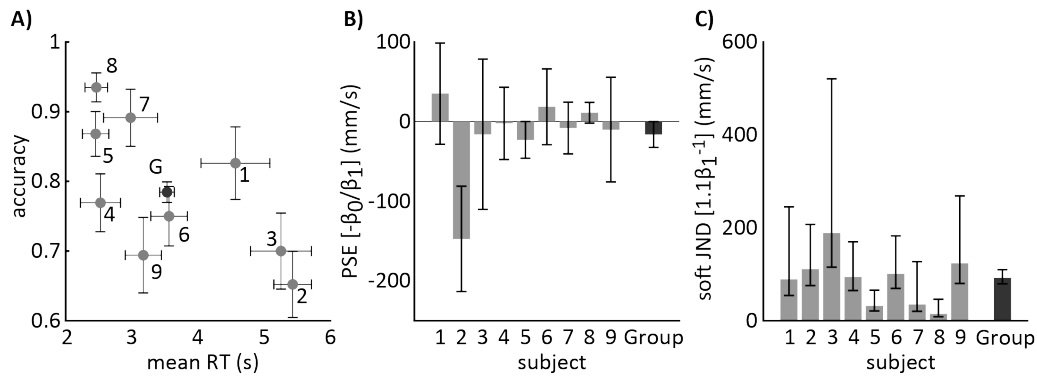


Figure 3: Individual variability in belt speed difference perception. **A)** mean reaction times (\pm standard error) vs. mean accuracy for each subject (numbered) and the whole Group across all stimulus sizes. **B)** estimates of point of subjective equality (PSE, thick bars) and 95% confidence interval (CI, errorbars) from logistic regression models. Best estimate and approximate confidence intervals are propagated from β_0 and β_1 estimates (PSE= $-\beta_0/\beta_1$) presuming fixed β_1 to its maximum likelihood value. **C)** just noticeable difference (JND) estimate with 95% CI (errorbars). Values are computed as $1.1/\beta_1$. Confidence intervals are computed by applying the same transformation to the edges of the CI of β_1 . This results in skewed CIs.

411 ranging from approximately 2.5 to 5.5 s. For PSE and the JND determi-
 412 nation, choices were modeled for each individual considering the same two
 413 factors as in the group level (stimulus size and intercept, Figure 2A, gray
 414 lines). A representation of individual parameter estimates is shown in Fig-
 415 ure 3 (Panel B: PSE; Panel C: JND). Only one of the subjects showed a
 416 significant bias (PSE, Subject 2), which we believe was driving the intercept
 417 term in the group level regression. A large range of values is observed for
 418 the JND estimate obtained from this model, implying that some subjects
 419 have much sharper discrimination curves than others. The observed thresh-
 420 old range was 18 to 191 mm/s, with an average of 96 ± 59 mm/s (mean \pm
 421 standard deviation).

422 **The drift-diffusion model can explain accuracy and reaction**
423 **times in this task.** This model represents the decision-making process
424 in the 2AFC task as the random walk of a variable quantifying the evidence
425 accumulated over time. A choice is made once the evidence exceeds some
426 pre-determined values (see Figure 4, top). The drift-diffusion model (DDM)
427 could adequately fit experimental results. Figure 4 shows group-averaged
428 data along with group-averaged model predictions. Model parameters were
429 fit to each individual separately, assuming a linear relation between stimuli
430 (belt speed difference) and drift rate.

431 We note that a DDM with fixed noise could be fitted to adequately explain
432 accuracy or reaction times (RT), but it was not possible to describe both
433 outcomes simultaneously (results not shown). For the parameters that best
434 fit RT, the expected subject accuracy was higher than the empirical one.
435 Conversely for the parameters that best fit accuracy the expected RT curve
436 was flatter than the observations. We conclude that the DDM as presented
437 is sufficient to characterize both subjects' accuracy and reaction times if the
438 model's noise, or diffusion rate, is allowed to be stimulus-dependent.

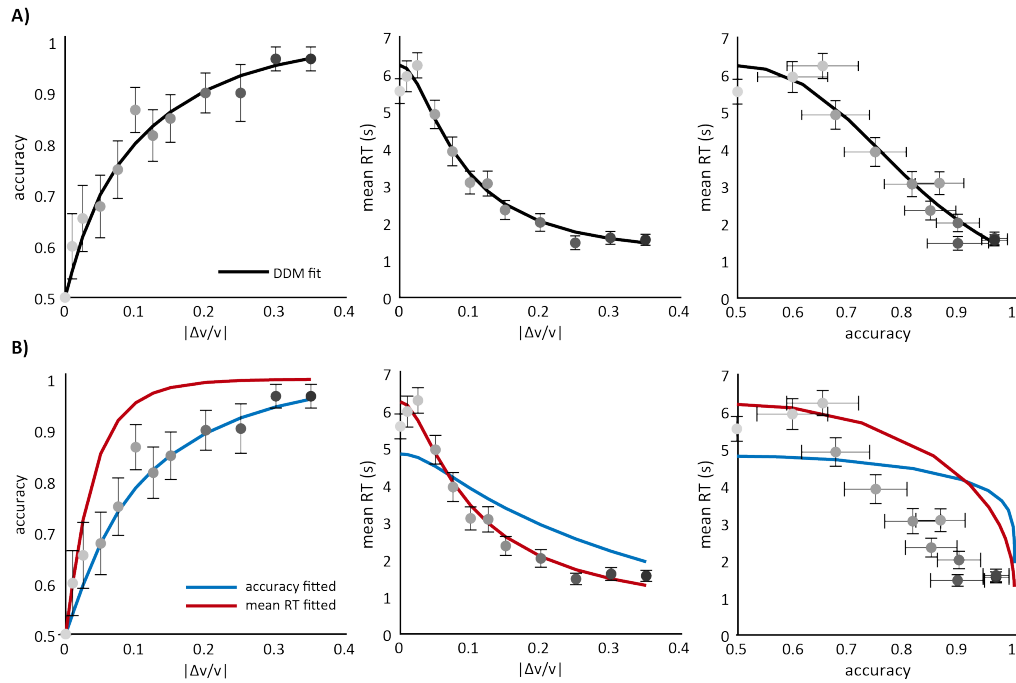


Figure 4: Drift diffusion models with fixed noise are insufficient to fit 2AFC task data. **A) Left panel:** accuracy vs. absolute stimulus size. Black line represents the model fit, circles indicate experimental data (group average \pm standard error). **Middle Panel:** mean reaction time (RT) vs. absolute stimulus size. **Right Panel:** mean RT vs. accuracy. **B)** The model with fixed noise could be fitted to adequately explain accuracy (Blue line) or mean reaction time (Red line). **Left Panel:** accuracy vs. absolute stimulus size. **Middle Panel:** mean reaction time (RT) vs. absolute stimulus size. **Right Panel:** mean RT vs. accuracy.

439 5 Discussion

440 5.1 Quantification of Belt Speed Difference Sensation 441 Through a 2AFC Task

442 This work presents a rigorous assessment of subjects' sensation of differences
443 in belt speeds on a split-belt treadmill. Specifically, we present results on
444 the accuracy of subjects in identifying which belt was moving faster than
445 the other as a function of the magnitude of belt speed differences. Differ-
446 ent from prior studies on the topic that relied on variants of a yes/no task
447 (Lauzière et al., 2014; Hoogkamer et al., 2015; Wutzke et al., 2015), we uti-
448 lized a two-alternative forced-choice (2AFC). Yes/no task responses depend
449 on both available sensory information and of subject's valuation of the two
450 alternative responses. For example, when asked if belts are moving at the
451 same speed, subjects favor one response when in doubt, while others may
452 prefer the alternative answer. However, none of the prior studies explicitly
453 account for this decision criteria. The 2AFC presents alternative responses
454 in a symmetric design, so differences in value between the alternatives can be
455 assumed to be nonexistent. This allows for inference about the sensory in-
456 formation used to arrive at a response without needing to consider subjects'
457 decision criteria, making 2AFC the preferred task when the objective is to
458 assess sensory information and processes (Green and Swets, 1966). Conse-
459 quently, we believe the results from our work represent the most accurate
460 report to date on healthy young persons' sensation on this task.

461 **5.2 Middle Point JNDs are a More Meaningful Metric** 462 **of Sensitivity than Null Hypothesis JNDs**

463 The just noticeable difference (JND) is a useful metric to summarize sensa-
464 tion of a particular physical stimulus. Three prior reports on sensation of
465 belt speed differences in split-belt walking explicitly estimate JNDs in this
466 context (Lauzière et al., 2014; Hoogkamer et al., 2015; Wutzke et al., 2015).
467 However, these studies are not always explicit about what they define as the
468 JND, which makes interpretation of experimental results harder. In this work
469 we adopted and quantified our results through one commonly used definition:
470 the middle point threshold. An alternative definition is the one given by the
471 null hypothesis threshold (Goldstein, 2009). Both definitions (middle point
472 threshold and null hypothesis) quantify different notions of what a JND is,
473 and either or both may be adopted to characterize probabilistic relation be-
474 tween detection and stimulus size. However, we believe that the middle point
475 JND quantification is more meaningful when characterizing sensitivity.

476 The null hypothesis JND can be qualitatively defined as the point below
477 which stimuli are not reflected in sensory information that is available to
478 subjects' response mechanisms. Consequently, belt speed differences below
479 this threshold should result in chance-level (50%) choices. Here we found that
480 subjects, when taken as a group, were able to detect above chance levels, belt
481 speed differences as small as 25 mm/s when the mean belt speed was 1.05
482 m/s. This corresponds to a 2.4% Weber fraction. Subjects' accuracy in
483 determining belt speed differences of 10 mm/s (the smallest tested here) was
484 60%, but this value was not significantly different from chance. These results
485 suggest that the null hypothesis threshold in this context is certainly below

486 2.4%, and possibly lies between 1 and 2.4%.

487 While null hypothesis JND may make intuitive sense, it is impossible to
488 positively prove its existence. Any negative findings (i.e., chance-level accu-
489 racy for some non-zero stimuli) can be explained away as a lack of statistical
490 power to assess true accuracy. In our case, if the true underlying accuracy for
491 10 mm/s differences is 60% (as estimated here), then experimental determi-
492 nation of this value to be above chance levels with 80% power and a 5% type
493 I error rate would require over 150 samples. Hence, one possible interpreta-
494 tion of our results is that the experiment was underpowered to detect a 60%
495 accuracy rate for 10 mm/s differences in belt speeds. Of course, this problem
496 can be controlled if a definition exists of what is the minimum accuracy level
497 that we deem to be meaningfully above chance (e.g., defining that accuracy
498 below 55 % is not meaningfully above chance even if it may be statistically
499 significant for a large enough sample size). Thus, we believe estimation of
500 accuracy rates for any particular stimulus size (e.g., defining estimates and
501 confidence intervals of accuracy for any given belt speed difference) is a more
502 meaningful way to describe sensation than ascertaining the existence of a
503 precise cutoff point.

504 The middle point JND corresponds to the point at which subjects are
505 more likely than not to correctly identify belt speed differences. In our task
506 this point corresponds naturally to the 75% accuracy point in the accuracy
507 vs. stimulus strength. This definition has the advantage of being a good
508 descriptor of perceptual sensitivity, regardless of whether thresholds in the
509 null hypothesis sense exist. Using this metric we found a threshold of 9.1%
510 for group-averaged accuracy data. Further, we were able to quantify this

511 for individual subjects, and we found a wide range of sensitivities at the
512 individual level, with a population average also of $9.1\% \pm 5.6\%$ (mean \pm
513 standard deviation). Although the results from the psychometric fits suggest
514 a threshold of 9.1%, the grouped data suggests a lower threshold (Figure
515 2B, top panel). Note that the grouped data estimated over the 75 mm/s
516 stimulus has an accuracy of exactly 75%, suggesting that the group-level JND
517 is at most 75 mm/s and the psychometric fit is overestimating the threshold.
518 Notably, these estimates are well below previously reported perceptual belt
519 speed thresholds.

520 Because of differences in threshold definitions, along with the previously
521 mentioned methodological differences, comparisons between studies should
522 be made cautiously. A summary of prior reports on split-belt treadmill JNDs
523 is given in Table 1. Wutzke et al. (2015) explicitly use a middle point defini-
524 tion. The two other reports (Lauzière et al., 2014; Hoogkamer et al., 2015)
525 implicitly adopt a non-probabilistic approach, quantifying thresholds from a
526 single, or at most two, independent measurements. This approach appears
527 more consistent with a null hypothesis framework, in which stimuli fall in
528 either the undetectable (chance level accuracy) or the detectable (100% ac-
529 curacy) categories. Whatever the definition, our results reflect higher sensory
530 acuity from healthy subjects than had previously been implied. This is con-
531 sistent with prior observations that subjects require less stimulus information
532 to make a decision in forced-choice tasks (Ehrenstein and Ehrenstein, 1999).

533 Throughout this report we have presented metrics of sensitivity to belt
534 speed differences as a Weber fraction, or % of mean belt speed. This normal-
535 ization procedure is based on the finding, across several sensory modalities,

Authors	Population	Procedure	JND
Lauzière et al. (2014)	healthy elderly	y/n ascending limit	12.8%
Lauzière et al. (2014)	healthy elderly	y/n descending limit	16.2%
Wutzke et al. (2015)	chronic poststroke	y/n staircase	26%
Hoogkamer et al. (2015)	healthy young	y/n ascending limit	13%
Hoogkamer et al. (2015)	cerebellar	y/n ascending limit	17.6%
this study	healthy young	2AFC	< 9%

Table 1: Summary of split-belt perceptual thresholds reported in the literature, presented as a fraction of mean belt speed as a normalization procedure (Weber fraction, $w = \Delta v/\bar{v}$).

536 that the ability to perceive a difference between two stimuli scales linearly
537 with stimulus magnitude (Goldstein, 2009). This notion was not directly
538 tested here, but prior reports on this context offer support for this normal-
539 ization. Specifically, thresholds corresponding to equivalent Weber fractions
540 have been found in studies using varied walking speeds (Lauzière et al., 2014).
541 Verification of this relation is left for future studies.

542 **5.3 The Drift-Diffusion Model is Sufficient to Describe** 543 **Choices and Reaction Times in the 2AFC Task**

544 Drift-diffusion models (DDMs) have been successfully used to describe a wide
545 range of perceptual tasks (Ratcliff, 1978; Gold and Shadlen, 2007). One
546 of the appeals of DDMs is that it offers a framework to relate perceptual
547 outcomes (choice) with reaction times, which are not clearly related to the
548 task objective, through a computational description that depends on only a
549 few parameters. Further, the DDM can be interpreted as a formalization of
550 a process of statistical inference based on sequentially acquired information
551 Bogacz et al. (2006); Gold and Shadlen (2007). In our task DDMs with noise

552 levels that are stimulus-dependent were able to describe both accuracy or
553 reaction times as a function of stimulus strength. This validates the use of
554 DDMs in this type of perceptual task, which differs from prior applications
555 in that it is a movement task, and in which decisions are made over seconds
556 to tens of seconds, rather than in hundreds of milliseconds.

557 Recent work has suggested (Pardo-Vazquez et al., 2019) that this type
558 of decision-making models, when combined with experimental results made
559 under different stimuli combinations, can be informative about the neural
560 coding of said stimuli. Thus, it is of interest to understand if relations such
561 as Weber’s law hold in this context too.

562 Relatedly, it would be interesting to study the problem of optimal decision
563 barriers on this task as a function of a time vs. accuracy trade-off across
564 different stimuli. The study of these potential modifications to the model is
565 left for future studies.

566 **References**

- 567 Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J. D. (2006).
568 The physics of optimal decision making: A formal analysis of models of
569 performance in two-alternative forced-choice tasks. *Psychological Review*,
570 113(4):700–765.
- 571 Dietz, V., Zijlstra, W., and Duysens, J. (1994). Human neuronal interlimb
572 coordination during split-belt locomotion. *Experimental Brain Research*,
573 101(101):513–520.
- 574 Ehrenstein, W. and Ehrenstein, A. (1999). Psychophysical methods. In

- 575 *Modern Techniques in Neuroscience Research*, chapter 43, pages 1211–
576 1241. Springer, Berlin.
- 577 Finley, J. M., Long, A., Bastian, A. J., and Torres-Oviedo, G. (2015). Spatial
578 and temporal control contribute to step length asymmetry during split-belt
579 adaptation and hemiparetic gait neurorehabilitation. *Neurorehabilitation
580 and Neural Repair*, 29(8):786–795.
- 581 Gold, J. I. and Shadlen, M. N. (2007). The neural basis of decision making.
582 *Annual Review of Neuroscience*, 30:535–574.
- 583 Goldstein, E. B. (2009). *Encyclopedia of Perception*. SAGE Publications,
584 Inc, 9th edition.
- 585 Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psy-
586 chophysics*. John Wiley and Sons.
- 587 Henmon, V. (1911). The relation of the time of a judgment to its accuracy.
588 *Psychological Review*, 18(3):186–201.
- 589 Hoogkamer, W., Bruijn, S. M., Potocanac, Z., Van Calenbergh, F., Swin-
590 nen, S. P., and Duysens, J. (2015). Gait asymmetry during early split-belt
591 walking is related to perception of belt speed difference. *Journal of Neu-
592 rophysiology*, 114:1705–1712.
- 593 Jensen, L., Prokop, T., and Dietz, V. (1998). Adaptational effects during
594 human split-belt walking: influence of afferent input. *Experimental Brain
595 Research*, 118(1):126–130.

- 596 Lauzière, S., Miéville, C., Duclos, C., Aissaoui, R., and Nadeau, S. (2014).
597 Perception threshold of locomotor symmetry while walking on a split-
598 belt treadmill in healthy elderly individuals. *Perceptual and motor skills*,
599 118(2):475–90.
- 600 Leech, K. A., Day, K. A., Roemmich, R. T., and Bastian, A. J. (2018).
601 Movement and perception recalibrate differently across multiple days of
602 locomotor learning. *J Neurophysiol*, 120:2130–2137.
- 603 Mawase, F., Shmuelof, L., Bar-Haim, S., and Karniel, A. (2014). Savings in
604 locomotor adaptation explained by changes in learning parameters follow-
605 ing initial adaptation. *Journal of Neurophysiology*, 111(7):1444–1454.
- 606 Ogawa, T., Kawashima, N., Ogata, T., and Nakazawa, K. (2014). Predic-
607 tive control of ankle stiffness at heel contact is a key element of locomotor
608 adaptation during split-belt treadmill walking in humans. *Journal of Neu-
609 rophysiology*, 111(4):722–732.
- 610 Pardo-Vazquez, J. L., Castiñeiras-de Saa, J. R., Valente, M., Damião, I.,
611 Costa, T., Vicente, M. I., Mendonça, A. G., Mainen, Z. F., and Renart, A.
612 (2019). The mechanistic foundation of Weber’s law. *Nature Neuroscience*,
613 22(9):1493–1502.
- 614 Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*,
615 85(2):59–108.
- 616 Reisman, D. S., Wityk, R., Silver, K., and Bastian, A. J. (2007). Locomotor
617 adaptation on a split-belt treadmill can improve walking symmetry post-
618 stroke. *Brain*, 130(7):1861–1872.

- 619 Sombric, C. J., Harker, H. M., Sparto, P. J., and Torres-oviedo, G. (2017).
620 Explicit Action Switching Interferes with the Context-Specificity of Motor
621 Memories in Older Adults. *Frontiers in Aging Neuroscience*, 9:40.
- 622 Statton, M. A., Vazquez, A., Morton, S. M., Vasudevan, E. V., and Bastian,
623 A. J. (2018). Making sense of cerebellar contributions to perceptual and
624 motor adaptation. *Cerebellum*, 17(2):111–121.
- 625 Torres-Oviedo, G. and Bastian, A. J. (2010). Seeing is believing: effects
626 of visual contextual cues on learning and transfer of locomotor adapta-
627 tion. *The Journal of neuroscience : the official journal of the Society for*
628 *Neuroscience*, 30(50):17015–17022.
- 629 Vazquez, A., Statton, M. a., Busgang, S. a., and Bastian, A. J. (2015). Split-
630 belt walking adaptation recalibrates sensorimotor estimates of leg speed,
631 but not position or force. *Journal of Neurophysiology*, page jn.00302.2015.
- 632 Wagenmakers, E. J., Van Der Maas, H. L. J., and Grasman, R. P. P. P.
633 (2007). An EZ-diffusion model for response time and accuracy. *Psycho-*
634 *nomic Bulletin and Review*, 14(1):3–22.
- 635 Wutzke, C. J., Faldowski, R. A., and Lewek, M. D. (2015). Individuals post-
636 stroke do not perceive their spatiotemporal gait asymmetries as abnormal.
637 *Physical Therapy*, 95(9):1244–1253.