1
2
3
4
5
6
7
8
9
10
11      **Custom-molded headcases have limited efficacy in reducing head motion for fMRI**
12
13                      Jolly, E[1*+]., Sadhukha, S[2*]., Chang, L.J.[1]
14
15                  [1]Computational Social Affective Neuroscience Laboratory
16              Department of Psychological and Brain Science, Dartmouth College
17                  [2]Department of Psychology, New York University
18
19
20      *Equal Contribution
21      +Corresponding Author
22
23      Corresponding author contact (eshin.jolly@dartmouth.edu)
24      6207 Moore Hall
25      Dartmouth College
26      Hanover, NH, 03755
27
28
29
30
31
32
33
34
35
36
37

## Abstract

Effectively minimizing head motion continues to be a challenge for the collection of functional magnetic resonance imaging (fMRI) data. The development of individual-specific custom molded headcases have been offered as a promising solution to minimizing motion during data collection, but to date, only a single published investigation into their efficacy exists in the literature. That study found headcases to be effective in reducing motion during short resting state fMRI scans (Power et al, 2019). In the present work, we examine the efficacy of these same headcases in reducing motion for a larger group of participants engaged in naturalistic scanning paradigms that consist of long movie-watching scans (~20-45min), as well as speaking aloud inside the MRI. Unlike previous work, we find no reliable reduction in head motion during movie viewing when comparing participants with headcases to those who were simply situated with foam pillows or foam pillows with medical tape. Surprisingly, we also find that for those wearing headcases, head motion is *worse* while talking relative to those situated with just foam pillows. These differences appear to be driven by large brief rotations of the head as well as translations in the z-plane as participants speak. Smaller, constant head movements appear equivalent with or without headcases. The largest reductions in head motion are observable when participants were situated with both foam pillows and medical tape, consistent with recent work by Kraus and colleagues (2019). Altogether, this work suggests that in a non-clinical, non-developmental population, custom-molded headcases may provide limited efficacy in reducing head motion beyond existing tools available to researchers. We hope this work can help improve the quality of custom headcases, motivate the investigation of additional solutions, as well as help researchers make more informed decisions about their data acquisition procedures.

## Keywords

fMRI, head-motion, framewise-displacement, naturalistic-paradigms, head-mold

## 1. Introduction

Head motion is a pernicious problem in functional magnetic resonance neuroimaging (fMRI) research and has been estimated to account for 30-90% of the total signal variance (Friston et al., 1996). Notably, motion during scans produces highly variable signal disruptions that dramatically change the readout of the global and voxel signals (Power et al., 2012; Satterthwaite et al., 2013; Van Dijk et al., 2012). Head motion is often exacerbated in developmental and clinical populations (Satterthwaite et al., 2012; Vanderwal et al., 2015) and makes it difficult to estimate task-specific activations when motion is correlated with stimulus onsets (Bullmore et al., 1999). Functional connectivity analyses (fcMRI) are particularly susceptible to head motion, which introduces spurious but systematic correlations across brain regions (Power et al., 2012; Van Dijk et al., 2012; Yan et al., 2013). Spurious correlations demonstrate regional variability and are often more pronounced in prefrontal areas including the default and prefrontal networks (Van Dijk et al., 2012; Yan et al., 2013).

A large body of work continues to investigate various post-acquisition preprocessing and analytic strategies to mitigate the effects of head motion on the fMRI signal, including but not limited to: ICA-based nuisance removal, PCA-based nuisance regression, voxel-specific realignment regression (including Volterra expansion), scrubbing, white-matter and cerebrospinal fluid space nuisance regression, and global signal regression (Behzadi et al., 2007; Friston et al., 1996; Hallquist et al., 2013; Jo et al., 2013; Mowinckel et al., 2012; Muschelli et al., 2014; Power et al., 2012; Satterthwaite et al., 2012; Siegel et al., 2014; Tyszka et al., 2014; Yan et al., 2013) (for a review see: (Power et al., 2015). However, another class of solutions involves *minimizing* head motion during the acquisition of brain volumes. While some technical solutions like prospective acquisition correction (Thesen et al., 2000) have been used to correct motion in near real-time, more common solutions involve situating individuals within the scanner in a way that restricts movement of the head in the first place. Such techniques include the use of foam head-coil stabilizers and a bite-bar, thermoplastic masks over the nose and brow, general foam padding packed into the head-coil, visual feedback systems that allow participants to adjust their head position during a study, and tactile feedback using medical tape (Bettinardi et al., 1991; Fitzsimmons et al., 1997; Krause et al., 2019; Menon et al., 1997; Power et al., 2019; Thulborn, 1999; Zaitsev et al., 2015). Other techniques include training, mock-scanning, and using specialized movie stimuli to improve compliance in developmental and clinical populations (Vanderwal et al., 2019). While these approaches have proved promising, they can also make collecting data more logistically challenging and can increase participant discomfort, resulting in poor widespread adoption except in the case of developmental or clinical populations (Zaitsev et al., 2015).

A recent novel solution involves the use of custom-molded head stabilizers ("headcases"), developed on a per-individual basis and conformant to an individual's unique anatomy. These custom head molds are milled from rigid styrofoam and distributed by the commercial company Caseforge (https://caseforge.co). Molds are produced from 3D optical scans of each participant's head, providing a custom fit that accounts for the shape of each individual's skull,

neck, and facial structure. Taking such factors into account has been claimed to prevent motion and more precisely position participants within a scanner, while also increasing comfort. While advertised as a promising alternative to previous approaches for reducing motion during acquisition, to date, only a single systematic investigation of the efficacy of these claims has been published in the literature (Power et al., 2019). One additional single participant dataset with and without headcases is also publicly available, but has been primarily used to investigate respiratory oscillations during multiband acquisition (Etzel & Braver, 2018). In their investigation, Power et al (2019) found that the use of headcases reduced motion during brief (4.8 min) resting state fMRI (rsfMRI) acquisitions in a group of 14 participants aged 7-28 years old. Their primary findings describe how headcases (1) decreased head motion in both rotational and translational axes, (2) reduced the fraction of a scan with large motions measured using mean and median Framewise Displacement (a composite measure of head motion calculated from the realignment of functional volumes (Power et al., 2012, 2015)), and (3) reduced the size of the small, constant motions throughout the scan.

In the present work we test the efficacy of these same custom-molded headcases in the context of "naturalistic" experimental designs that involve movie watching and talking aloud in the scanner environment. Many recent studies have utilized this "task-free" approach to probe neural representations and cognition because it can capture a larger degree of individual variability in cognitive and neural responding (Vanderwal et al., 2019). This approach also provides researchers with an opportunity to utilize rich datasets to ask a variety of questions, free of the constraints that come with pre-committing to a particular experimental design (Kriegeskorte et al., 2008). However, many of these studies often involve much longer acquisition times ranging from ~15 minute continuous runs e.g. (Haxby et al., 2011) up to ~45 minutes e.g. (Chang et al., 2018) providing more opportunities for participants to move (Meissner et al., 2019). Some studies additionally ask participants to speak aloud during scanning, certainly exacerbating head motion, as participants must enunciate organically, while lying as still as possible (Baldassano et al., 2017; Chen et al., 2017; Silbert et al., 2014; Stephens et al., 2010; Zadbood et al., 2017). These designs provide an excellent opportunity to examine the efficacy of custom headcases in reducing motion under more demanding situations that increase the likelihood of movement.

Therefore, the present work builds upon the examination by Power et al (2019) in several key ways: a) we compare movement from datasets representative of recent naturalistic experiments using much longer acquisitions (~45 continuous run); b) we report between group comparisons with larger sample sizes ($N$ = 26-35 vs $N$ = 14); c) we take advantage of between group comparisons that are matched on nearly every acquisition feature (i.e. scanner site, parameters, stimulus, etc.) or experimental task (i.e. active verbal recall of a previously watched movie) when matching acquisition features is not possible; d) compare movement from datasets in which participants are speaking aloud, providing a more rigorous test of the performance of headcases under more demanding scenarios. For consistency and direct comparison, we utilize the same approach as Power (2019), focusing primarily on Framewise Displacement as a global metric for head motion (Power et al., 2012, 2015), rather than changes in signal variability which

4

167 can be driven by stimulus content or participant specific idiosyncrasies (e.g. attention,
168 self-relevance) (Finn et al., 2019). We also examine individual translational and rotational
169 motion axes when participants talk aloud, to better identify the directions in which head motion
170 is exacerbated.
171
172 Our principal analyses focus on comparisons of the utility of custom-designed headcases while
173 participants *view* a single episode of a television show or *talk* aloud about the narrative of that
174 television show. Viewing comparisons comprise two datasets collected in an identical fashion at
175 the same site, from the same population, viewing the same episode either with or without
176 headcases. These comparisons also include an additional dataset collected in a similar fashion
177 albeit at a different site, from a different but comparable population, using different acquisition
178 parameters without headcases. This particular dataset served as an additional control to ensure
179 that any differences (or lack thereof) between the two similarly acquired datasets were not
180 driven by site, stimulus, or population idiosyncrasies. Talking comparisons comprise two of the
181 three viewing datasets which were collected at different sites, but involved the same
182 experimental task: freely recalling aloud the narrative of each respective television episode with
183 minimal time constraints.
184
185 **2. Methods**
186
187 All reported analyses are comprised of observations from three different datasets. Datasets 1
188 and 3 come from previously published studies (Chang et al., 2018) and (Chen et al., 2017). A
189 more detailed description of the data collection and preprocessing procedures are available in
190 those initial publications, but we provide an abbreviated summary of the methods here.
191 Separate manuscripts using Dataset 2 are forthcoming, but the subset of data used in the
192 current manuscript was collected in a fashion identical to Dataset 1 with the addition of
193 custom-molded headcases manufactured by Caseforge. Datasets 1 and 2 were collected at
194 Dartmouth College while Dataset 3 was collected at Princeton University. All participants
195 provided informed written consent in accordance with the experimental guidelines set by their
196 respective institutions: the Committee for the Protection of Human Subjects at Dartmouth
197 College and the Institutional Review Board at Princeton University.
198
199 **2.1 Dataset 1 (FNL no-headcase)**
200
201 2.1.1 *Participants & Procedure*
202 Thirty-five ($M_{age}$ = 19.0; $SD_{age}$ = 1.07*;* 26 female) Dartmouth College undergraduate students
203 were recruited from introductory psychology and neuroscience courses, participating for either
204 monetary compensation ($20/hr) or for partial course credit. Each participant watched one
205 episode of the television drama *Friday Night Lights* (FNL) while undergoing one continuous run
206 of fMRI. Audio was delivered using MR compatible in-ear headphones (Sensimetrics S14
207 https://www.sens.com/products/model-s14/). Participants were situated in the scanner with foam
208 pillows and medical tape attached to their foreheads which provided tactile feedback (Krause et

5

209 al., 2019) regarding head motion. Participants were also carefully instructed to lie as still as
210 possible. For more details, see methods for Study 2 in (Chang et al., 2018).

211

212 ### 2.1.2 *Imaging Acquisition*
213 Data were acquired at the Dartmouth Brain Imaging Center (DBIC) on a 3T Siemens Magnetom
214 Prisma scanner (Siemens, Erlangen, Germany) with a 32-channel phased-array head coil. Raw
215 DICOM images were converted to NIfTI images and stored in the brain imaging data structure
216 (BIDS) format using ReproIn from the ReproNim framework (K. J. Gorgolewski et al., 2016;
217 Visconti di Oleggio Castello et al., 2018). Functional blood-oxygenation-level-dependent (BOLD)
218 images were acquired in an interleaved fashion using gradient-echo echo-planar imaging with
219 pre-scan normalization, fat suppression, an in-plane acceleration factor of two (i.e. GRAPPA 2),
220 and no multiband (i.e. simultaneous multi-slice; SMS) acceleration: TR/TE: 2000/25ms, flip
221 angle = 75°, resolution = 3mm$^3$ isotropic voxels, matrix size = 80 x 80, FOV = 240 x 240mm$^2$ ,
222 40 axial slices with full brain coverage and no gap, anterior-posterior phase encoding.
223 Functional images were acquired in a single continuous run of 45.47 minutes (1364 TRs) which
224 began and ended with 5 TRs of fixation.

225

226 ## 2.2 Dataset 2 (FNL with headcase)

227

228 ### 2.2.1 *Participants*
229 Thirty-six ($M_{age}$ = 22.77; $SD_{age}$ = 4.73; 27 female) Dartmouth College undergraduate and
230 graduate students were enrolled for a three-part study, participating for either monetary
231 compensation ($20/hr) or for partial course credit. All reported data and analyses come from a
232 subset of part-one and part-three of this study. One participant rescinded their desire to
233 participate half-way through the first session and was consequently dropped from the dataset
234 entirely. A total of 7 subjects were excluded from all reported analyses due to issues with their
235 customized headcases: 4 participants' reported extreme discomfort with their headcases during
236 the first session, which resulted in no headcase use in subsequent sessions; 3 participants used
237 only the front or back of their headcases along with additional foam padding due to discomfort.
238 Two additional participants in this sample did not use headcases at all, but were situated in the
239 scanner using foam pillows and medical tape as in Dataset 1. This resulted in a total of 26
240 participants with headcases and 2 without headcases. These two participants were combined
241 with participants from Dataset 1 for all *viewing* comparisons reported below, but were not
242 included in any *talking* comparisons.

243

244 ### 2.2.2 *Procedure*
245 Across three experimental sessions that took place within approximately one week, participants
246 watched the first four episodes of the television show *Friday Night Lights* and performed several
247 memory tasks that involved talking aloud while undergoing fMRI. All reported analyses consist
248 of motion estimates while *viewing* the first episode during session one, and *talking* about all four
249 episodes during a spoken recall task in session three. This recall task was similar in nature to
250 the recall task used by Chen and colleagues (2017), in Dataset 3 (see below). Participants were
251 asked to recall aloud the narrative events of all four episodes they had previously seen. They

252 were given 1 minute to plan their responses and were asked to try to speak for a minimum of 10
253 minutes and a maximum of 30 minutes. This task was manually ended by experimenters when
254 individuals verbally indicated they were finished or automatically ended when the maximum
255 recall time elapsed. Audio recordings were acquired using a MR-compatible microphone and
256 recording system (Optoacoustic FOMRI III+
257 http://www.optoacoustics.com/medical/fomri-iii/features).

258

259 2.2.3 *Headcase production*
260 Prior to coming in for the multi-part study, participants were asked to visit the lab to have their
261 heads scanned with a handheld 3D scanner purchased from the CaseForge company. 360°
262 scans of each participant's head were acquired using procedures provided by CaseForge,
263 which were identical to those reported by Power et al (2019). Participants wore a swim cap
264 while being 3D photographed, as hair shape interfered with the quality and fit of the resulting
265 case molds. Images were uploaded to the Caseforge website which verified the quality of scans
266 and subsequently shipped a two piece customized styrofoam mold consisting of a front and
267 back half. Headcases were utilized in lieu of any additional padding within the head coil of the
268 MRI machine during acquisition. As in Dataset 1, participants were instructed to lie as still as
269 possible, particularly when speaking aloud. All headcases used in this dataset were
270 manufactured in late 2017 through August 2018.

271

272 2.2.4 *Imaging Acquisition*
273 Data were acquired at the Dartmouth Brain Imaging Center (DBIC) on a 3T Siemens Magnetom
274 Prisma scanner (Siemens, Erlangen, Germany) with a 32-channel phased-array head coil in a
275 manner identical to Dataset 1. Reported analyses come from a single continuous *viewing* run of
276 45.47 minutes (1364 TRs) which began and ended with 5 TRs of fixation and a variably ranged
277 *talking* run ($M_{Minutes}$ = 20.09; $SD_{Minutes}$ = 6.52; $Min_{Minutes}$ = 12.1; $Max_{Minutes}$ = 31.2) which began with
278 5 TRs of fixation and ended with 15 TRs of fixation.

279

280 **2.3 Dataset 3 (Sherlock)**

281

282 2.3.1 *Participants & Procedure*
283 Twenty-two participants ($M_{age}$ = 20.8, 10 female) from the Princeton community were recruited
284 for monetary compensation ($20/hr). Of this sample, 17 met exclusionary criteria in the
285 published sample by Chen et al (2017) and thus were used in all reported analyses. We direct
286 readers to the aforementioned manuscript for full procedural details, but in brief: participants
287 watched a 48-min segment of the BBC television series *Sherlock* and subsequently verbally
288 recalled the narrative of the show aloud while undergoing fMRI. During the recall task
289 participants were instructed to talk for a minimum of 10 minutes, and were allowed to talk for as
290 long as they wished. Experimenters manually ended the scanning run during the recall task
291 based on verbal indication from participants. Participants were situated with foam padding and
292 instructed to remain very still while viewing and speaking.

293

294

7

2.3.2 *Imaging Acquisition*

Data were collected on a 3T full-body scanner (Siemens Skyra) with a 20-channel head coil. Functional images were acquired using a T2*-weighted echo-planar imaging (EPI) pulse sequence (TR 1500 ms, TE 28 ms, flip angle = 64°, whole-brain coverage 27 slices of 4 mm thickness, in-plane resolution 3 × 3 mm2, FOV 192 × 192 mm2), with ascending interleaved acquisition. Reported analyses come from two *viewing* runs of 23 and 25 minutes long and a variably ranged *talking* run ($M_{Minutes}$ = 22.23; $SD_{Minutes}$ = 8.62; $Min_{Minutes}$ = 10.95; $Max_{Minutes}$ = 44.15).

**2.4 Motion Estimation and Comparison**

For all three datasets, head position and motion was estimated using the FSL tool MCFLIRT (Jenkinson et al., 2002) by realigning each volume to the mean volume of the run. This yielded three translation and three rotation estimates at each volume. These parameters were used to calculate Framewise Displacement (FD) using the approach in Power et al. (2012) and implemented in nipype (K. Gorgolewski et al., 2011). This metric reflects the summation of the absolute-value backwards-differences of each parameter. Angular rotation parameters were converted to arc displacement using a 50mm radius prior to summation (Power et al., 2012, 2015, 2019). For each participant, these motion time-series were used to calculate five summary statistics following the approach in Power et al (2019): mean and median displacement ($FD_{Mean}$, $FD_{Median}$), proportion of high-motion volumes that exceeded 0.3mm of displacement ($Spike_{Proportion}$), and mean and median displacement excluding high-motion volumes ($FD_{MeanFiltered}$, $FD_{MedianFiletered}$). In Dataset 3, because participants viewed the stimulus across two separate runs, motion estimates were first calculated and summarized *within run* and the average of each pair of summary statistics was used for all analyses.

In order to account for different numbers of individuals in each dataset, all group comparisons were performed using permuted independent-groups non-equal variance t-tests implemented in Pymer4 (Jolly, 2018). First, for each comparison, a t-statistic was computed using Scipy (Jones et al., 2001--). Then, group labels (i.e. with or without headcase) were randomly shuffled while retaining the original group sizes and a new t-statistic was computed. This procedure was repeated 5,000 times to generate a null distribution of t-statistics. P-values were calculated by computing the number of permutations that were greater than or equal to the original t-statistic with adjustment to avoid non-zero p-values (Phipson & Smyth, 2010). Bootstrapped 95% confidence intervals around the mean difference between groups for each metric were also computed in Pymer4 by resampling with replacement from each group 5000 times while preserving the original group sizes.

To further interrogate non-significant results, equivalence tests were performed using the two-one-sided-tests (TOST) procedure (Lakens et al., 2018; Schuirmann, 1981) implemented in the pingouin python statistics library (Vallat, 2018). This was performed to estimate whether any non-significant differences between groups fell within a predefined range of *practical*

8

338 *equivalence*. In other words, non-significant group differences in motion alone do not provide
339 information about practical differences between groups that may still be of interest to fMRI
340 researchers, as small differences may still be detectable with large enough sample sizes. We
341 chose the equivalence bounds of +/- 0.05mm for the *viewing* condition based upon the findings
342 from Power et al (2019), and a larger range of +/- 0.15mm for the *talking* condition given the
343 increased motion as a result of speaking aloud in the scanner. In the *viewing* condition,
344 comparisons were made between the FNL without headcases (Dataset 1) and FNL with
345 headcases (Dataset 2) samples, while in the *talking* condition, comparisons were made
346 between the Sherlock (Dataset 3) and FNL with headcases (Dataset 2) samples. For each
347 comparison, two parametric, one-tailed, independent-groups, non-equal variance t-tests were
348 performed by adding/subtracting the equivalence bounds to the mean of one sample prior to
349 running each comparison (Vallat, 2018). Consistent with prior conventions, the reported p-value
350 reflects the larger of these two one-sided tests (Lakens et al., 2018).

351

352 **3. Results**

353

354 3.1 *Lack of overall motion reduction with headcases while viewing a movie*

355

356 Based on previous findings (Power et al., 2019), we expected to see reliable reductions in head
357 motion when participants were fitted with custom headcases during fMRI scanning. However,
358 while viewing the same stimulus (episode 1 of *Friday Night Lights)*, collected on the same
359 scanner, at the same site, with the same acquisition parameters, we found no significant
360 difference in $FD_{Mean}$ (Fig 1 top row, left column, blue and pink bars). Participants with headcases
361 moved equivalent amounts on average ($M$ = 0.129; $SD$ = 0.102) compared to participants
362 situated with only foam pillows and medical tape ($M$ = 0.113; $SD$ = 0.045), $t$ = -0.758, $p$ = 0.494.
363 This was also true when comparing $FD_{Median}$, FNL with-case ($M$ = 0.078; $SD$ = 0.03), FNL
364 no-case ($M$ = 0.085; $SD$ = 0.029), $t$ = 0.857, $p$ = .393, and the proportion of volumes with motion
365 in excess of 0.3mm ($Spike_{Proportion}$) FNL with-case ($M$ = 0.06; $SD$ = 0.075), FNL no-case ($M$ =
366 0.045; $SD$ = 0.048), $t$ = -0.917, $p$ = .363 (Fig 1 middle and bottom rows, left column, blue and
367 pink bars).

368

369 Despite being matched on nearly every dimension, we sought to ensure that our non-headcase
370 sample exhibited motion typical of the range observed in similar naturalistic imaging studies,
371 thus ensuring a fair comparison to our headcase sample. To do so, we compared motion
372 estimates from our FNL non-headcase sample to those from a previously published dataset, in
373 which individuals watched the first 48 minutes of the crime drama *Sherlock* (Chen et al., 2017)
374 and were situated with foam pillows. Our non-headcase sample exhibited no significant
375 differences in $FD_{Mean}$ $t$ = -0.703, $p$ = .489, or $Spike_{Proportion}$ $t$ = 0.723, $p$ = .466, but did exhibit a
376 significantly lower $FD_{Median}$, $t$ = -2.061, $p$ = .045. This translated to a significantly lower $FD_{Median}$
377 for our *FNL headcase* sample relative to *Sherlock*, $t$ = 2.60, $p$ = .015. This suggests that motion
378 in our non-headcase sample was comparable (and even slightly lower) relative to a similar
379 existing non-headcase dataset. This also suggests that the lack of a significant difference

9

between our non-headcase and headcase samples was unlikely due to something particularly unique about this dataset (Table 1).

However, because a lack of statistically significant differences does not necessarily provide evidence for the null, we instead tried to better quantify these null results using equivalence testing (Schuirmann, 1981). Using the TOST procedure, we defined +/- 0.05mm as the *equivalence bounds* of our comparisons, i.e. a range of mean-differences in head motion that fMRI researchers may consider "statistically equivalent" in (Lakens et al., 2018). This range was set based upon the reported *within-participant* average improvement in $FD_{Mean}$ by Power et al (2019) as a result of headcase use. We found that the observed mean differences and bootstrapped 95% confidence intervals for all motion metrics (Table 1) between headcase and non-headcase participants participants fell within this equivalence range: $FD_{Mean}$ $p$ = 0.061[1]; $FD_{Median}$ $p$ < .001 (Fig 3, left column, blue points). Together, these findings suggest that headcases provide limited efficacy in reducing overall head motion in longer, more arduous, scanning scenarios that present more opportunities for participants to move.

3.2 *Lack of small motion reduction with headcases while viewing a movie*

Following the approach of Power et al (2019), we repeated the previous analysis, this time excluding high-motion volumes per individual (FD > 0.3mm) prior to computing and comparing summary statistics. This analysis assesses the efficacy of headcases in reducing smaller, constant head movements by ignoring parts of each individual's scan that contain substantial motion. Our findings are largely similar to the previous results. We found no significant differences in $FD_{MeanFiltered}$ (Fig 2 top row, left column, blue and pink bars) or $FD_{MedianFiltered}$ when comparing participants with and without headcases viewing the same stimulus. The $FD_{Mean}$ of participants with headcases was equivalent on average ($M$ = 0.085; $SD$ = 0.026) compared to participants situated with only foam pillows and headtape ($M$ = 0.093; $SD$ = 0.025), $t$ = 1.186, $p$ = .244. This was also true of $FD_{Median}$, FNL with-case ($M$ = 0.073; $SD$ = 0.025), FNL no-case ($M$ = 0.081; $SD$ = 0.025), $t$ = 1.33, $p$ = .192. Equivalence tests using the same range as before also suggested that observed differences in $FD_{MeanFiltered}$ and $FD_{MedianFiltered}$ were of practical equivalence, $ps$ < .001 (Fig 3, left panel, red points).

Our control analyses comparing $FD_{MeanFiltered}$ and $FD_{MedianFiltered}$ between our non-headcase sample and the *Sherlock* sample of participants without headcases produced similar results. We observed no significant difference in $FD_{MeanFiltered}$, $t$ = -1.541, $p$ = .134, but did observe a difference in $FD_{MedianFiltered}$ $t$ = -2.368, $p$ = .022. This translated to a significant $FD_{MeanFiltered}$ difference between the *Sherlock* sample and *FNL headcase* sample $t$ = 2.43, $p$ = .022 as well as a significant $FD_{MedianFiltered}$ $t$ = 3.319, $p$ =.002 (Fig 2 bottom row, left column; Table 2).

---

[1] Reporting conventions for TOST results typically reflect the *higher* p-value of each one-sided test. In this case the higher value reflects the *lower bound* of -0.05mm which reflects a motion *increase* as a result of using headcases. Observed mean differences however, are fully contained within the *upper bound* of 0.05mm $p$ < 0.001, which reflects a motion *decrease* and is likely of more interest to fMRI researchers.
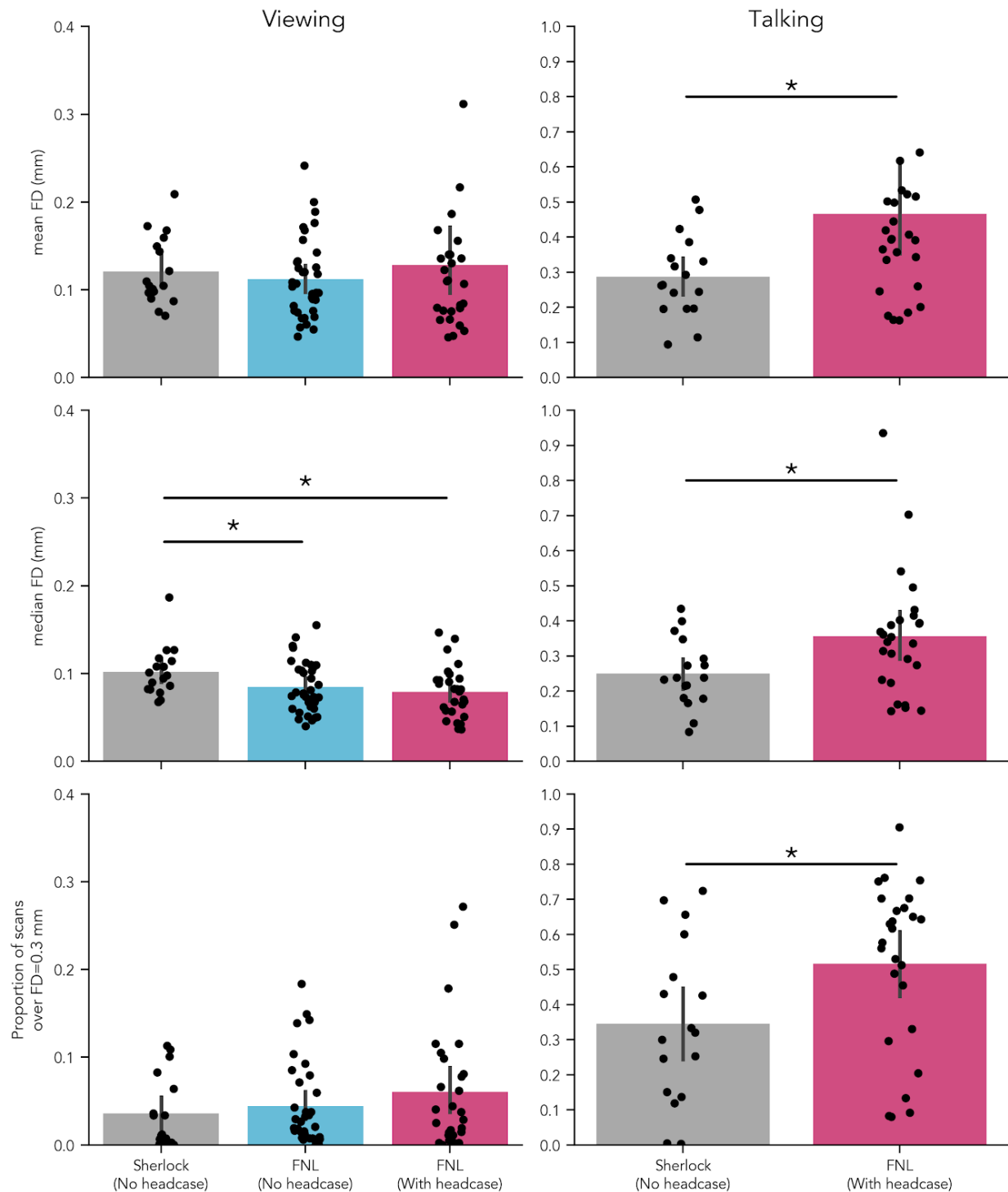
418



**Figure 1 | Headcase effects on Framewise Displacement when viewing and speaking inside the scanner**

419 Top row: average of participant mean FD; middle row: average of participant median FD; bottom row: proportion of
420 TRs in which FD exceeded 0.3mm. Blue and pink bars reflect two groups of participants who watched the same
421 stimulus (Friday Night Lights) under the same data collection procedures (e.g. parameters, scanner site) except for
422 the use of headcases. The grey bars reflect a group of subjects who watched a different stimulus (Sherlock) collected
423 at a different site with different acquisition parameters. Sherlock participants, however, performed the same talking
424 task inside the scanner albeit without headcases (right column). No significant improvement in mean or median FD or

proportion of high motion TRs was observed with the use of headcases during viewing between participants watching the same stimulus. However, a significant difference was observed between the median FD of the Sherlock sample and both no headcase and headcase wearing FNL samples. During talking however, all metrics suggested a significant increase in motion while wearing headcases.
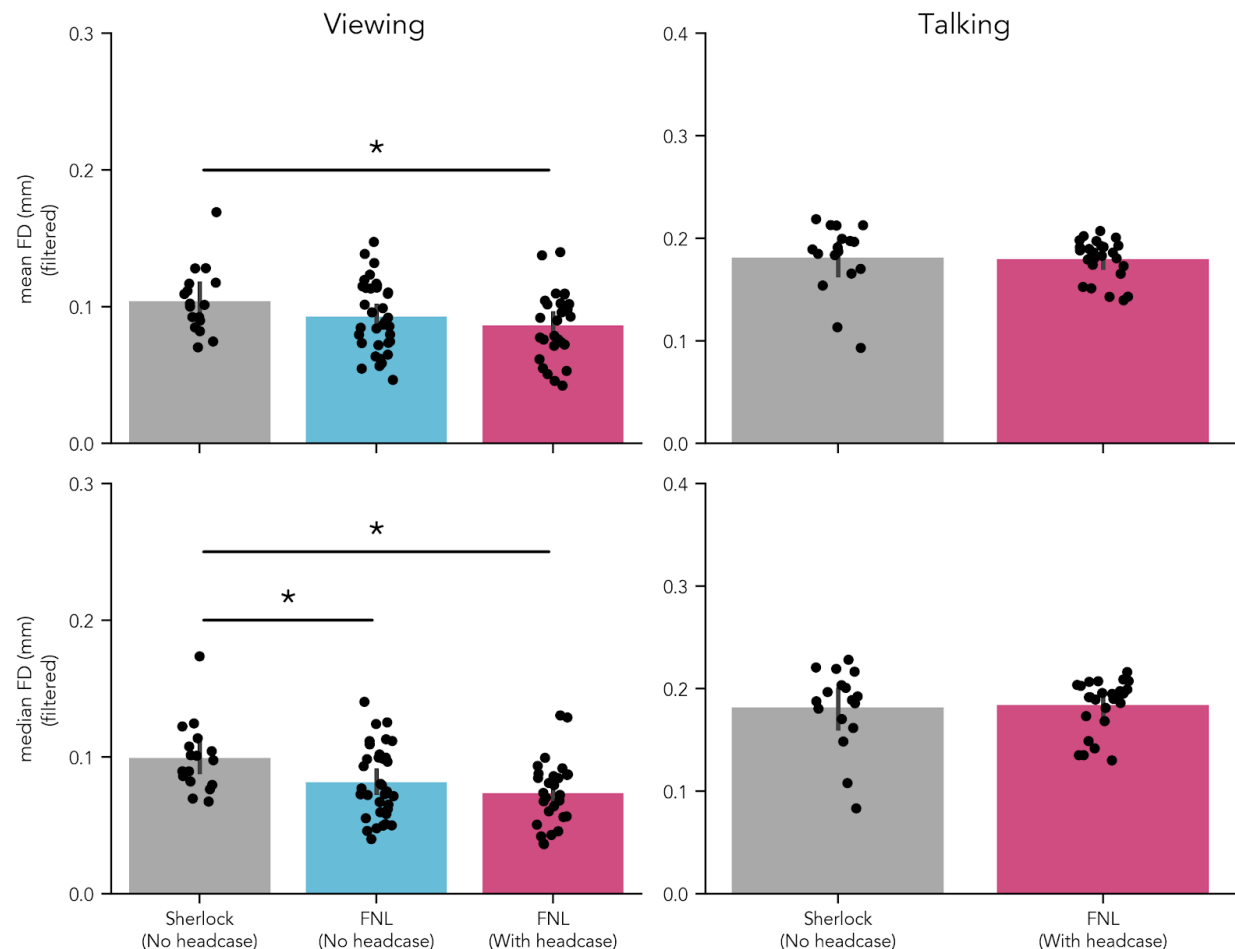


**Figure 2 | Headcase effects on Framewise Displacement excluding high motion volumes**
Mean and median FD differences after removing high motion TRs (FD > 0.3mm). Headcases demonstrated no detectable improvement in small head motions either during viewing or during talking. However, the Sherlock sample showed increased median FD during viewing relative to both FNL samples. Motion in the Sherlock sample was significantly higher than both FNL samples during viewing. Mean FD was higher in the Sherlock sample relative to the FNL sample with headcases.

### 3.3 *Lack of overall motion reduction with headcases while talking aloud*

While the previous analyses focused on a passive experimental "task" with a single long continuous run, our second set of analyses focused on an even more pernicious scenario for motion in neuroimaging - speaking during the scan. While several recent studies have utilized verbalization tasks in the scanner (Chen et al., 2017; Silbert et al., 2014; Stephens et al., 2010; Zadbood et al., 2017), none have compared the efficacy of custom headcases in mitigating

12

445 talking-induced head motion to our knowledge. We expected that even if headcases provided
446 minimal reduction of head motion during passive viewing, perhaps due to increased participant
447 compliance (Vanderwal et al., 2019), they *should* provide maximal benefit when participants
448 move their heads as a consequence of task demands (i.e. talking). However, to our surprise, we
449 did not find evidence supporting this hypothesis, and in some cases found the *opposite* result.
450 When comparing *Sherlock* participants without headcases to our *FNL headcase* participants in
451 the talking task, we found that participants wearing headcases exhibited *increased* motion
452 across all metrics[2] (Fig 1 right column). Participants wearing headcases exhibited higher $FD_{Mean}$
453 ($M$ = 0.467; $SD$ = 0.362) compared to *Sherlock* participants ($M$ = 0.287; $SD$ = 0.116), $t$ = -2.353,
454 $p$ = .009. They also exhibited higher $FD_{Median}$ ($M$ = 0.356; $SD$ = 0.177) relative to participants
455 without headcases ($M$ = 0.25; $SD$ = 0.097), $t$ = -2.534, $p$ = .013, as well as a larger proportion of
456 high-motion volumes ($Spike_{Proportion}$): with headcase ($M$ = 0.516; $SD$ = 0.237), without headcase
457 ($M$ = 0.345; $SD$ = 0.23), $t$ = -2.353, $p$ = .025. These results suggest that headcases lack the
458 efficacy to mitigate large head movement (e.g. > 0.3mm) observed in more pernicious scenarios
459 like speaking during a scan, despite the extra restriction they place on participants (Table 3).
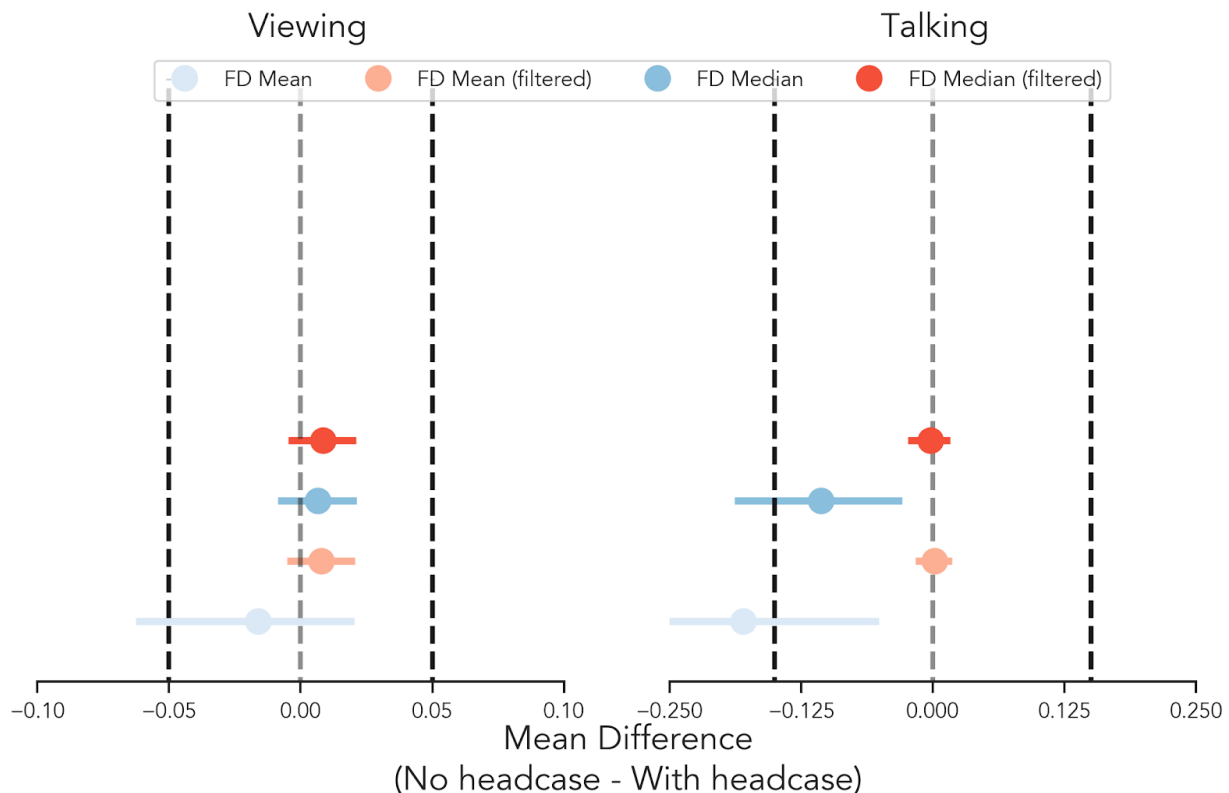460



**Figure 3 | Equivalence tests of motion estimates with and without headcases**
461 Plots depict whether mean differences between groups fall within bounds of practical value to fMRI researchers.
462 Mean differences within these bounds can be interpreted as "statistically equivalent" (Lakens et al., 2018). Darker
463 dashed lines indicate the upper and lower bounds of the equivalence tests, while the lighter dashed lines reflect a

---

[2] We reran these comparisons excluding two headcase participants who exhibited particularly high-levels of motion overall ($FD_{Mean}$ >= 1mm; data points not depicted in Figures), but found that headcase participants still exhibited significantly higher $FD_{mean}$ albeit equivalent levels of motion reflected in $FD_{Median}$ and $Spike_{Proportion}$ (see Supplementary Materials).

difference of 0. The upper positive bound (right dashed lines) in each plot tests whether head motion *decreased* by at least 0.05mm (viewing) or 0.15mm (talking) when wearing headcases. The lower negative bound (left dashed line) tests whether head motion *increased* by at least 0.05mm (viewing) or 0.15mm (talking). In all cases, motion statistics were significantly below the upper equivalence bound suggesting that observed mean reductions in motion while wearing headcases are unlikely to be as large as 0.05mm when viewing or 0.15mm when talking. Estimates that included high motion volumes during the talking condition (right blue points) also fell below the lower equivalence bound, suggesting that in some situations headcases may exacerbate motion by at least 0.15mm.

### 3.4 *Lack of small motion reduction with headcases while talking aloud*

We once again repeated the previous comparison after excluding high-motion volumes per individual to examine the efficacy of headcases on smaller head movements while talking. Headcases demonstrated no significant reduction in $FD_{MeanFiltered}$ between our sample ($M$ = 0.18; $SD$ = 0.019) and the *Sherlock* sample (M = 0.181; SD = 0.034) $t$ = 0.19, $p$ = .857. This was also true when examining $FD_{MedianFiltered}$: with head case ($M$ = 0.184; $SD$ = 0.025), without headcase ($M$ = 0.182; $SD$ = 0.039), $t$ = -0.222, $p$ = .829. To further quantify these null results (Fig 4 right column), we again performed equivalence testing using the TOST procedure, this time with larger equivalence bounds of +/- 0.15mm given the motion-inducing nature of the task. We found that for both $FD_{MeanFiltered}$ and $FD_{MedianFiletered}$, observed differences in motion fell within this range of practical equivalence (Fig 5 right column, red points), $ps$ < .001. These findings suggest that when specifically examining the reduction of smaller head movements induced while talking aloud, at best, headcases may be as efficacious as foam pillows alone (Table 3).

### 3.5 *Causes of increased motion while wearing headcases and talking*

Next, we explored what may have caused more motion while participants wearing headcases spoke aloud. We speculated that the interaction between lower jaw movements and head restriction may have paradoxically focused motion in the z-axis translation (moving head inward/outward parallel to the main axis of the scanner bore) and pitch-axis (nodding head up and down) rotation motion parameters. In other words, we hypothesized that because participants were largely restricted in every direction, but were freely moving their lower jaws, movements of the head may have been exacerbated in these planes as headcases provide no restriction of chin or lower jaw movement or tactile feedback. To test this hypothesis, we repeated our group comparison while participants talked *separately* for each axis of translation and rotation (Fig 4). Specifically, we compared displacement in the x, y, z, pitch, roll, and yaw axes using each participant's mean, median, and standard deviation displacement. Across all three summary statistics, we found significantly greater displacement in all rotation axes (pitch, roll, yaw), all $ps$ < 0.05 (Table 4). We also found marginally greater mean displacement in the z-axis, $p$ = 0.056, and significantly greater variability of displacement in both the x and z axes $ps$ < 0.05. These findings suggest that observed differences in overall $FD_{Mean}$, $FD_{Median}$ and $Spike_{Proportion}$ may have been driven by participants rotating their heads during talking, despite being constrained by headcases.
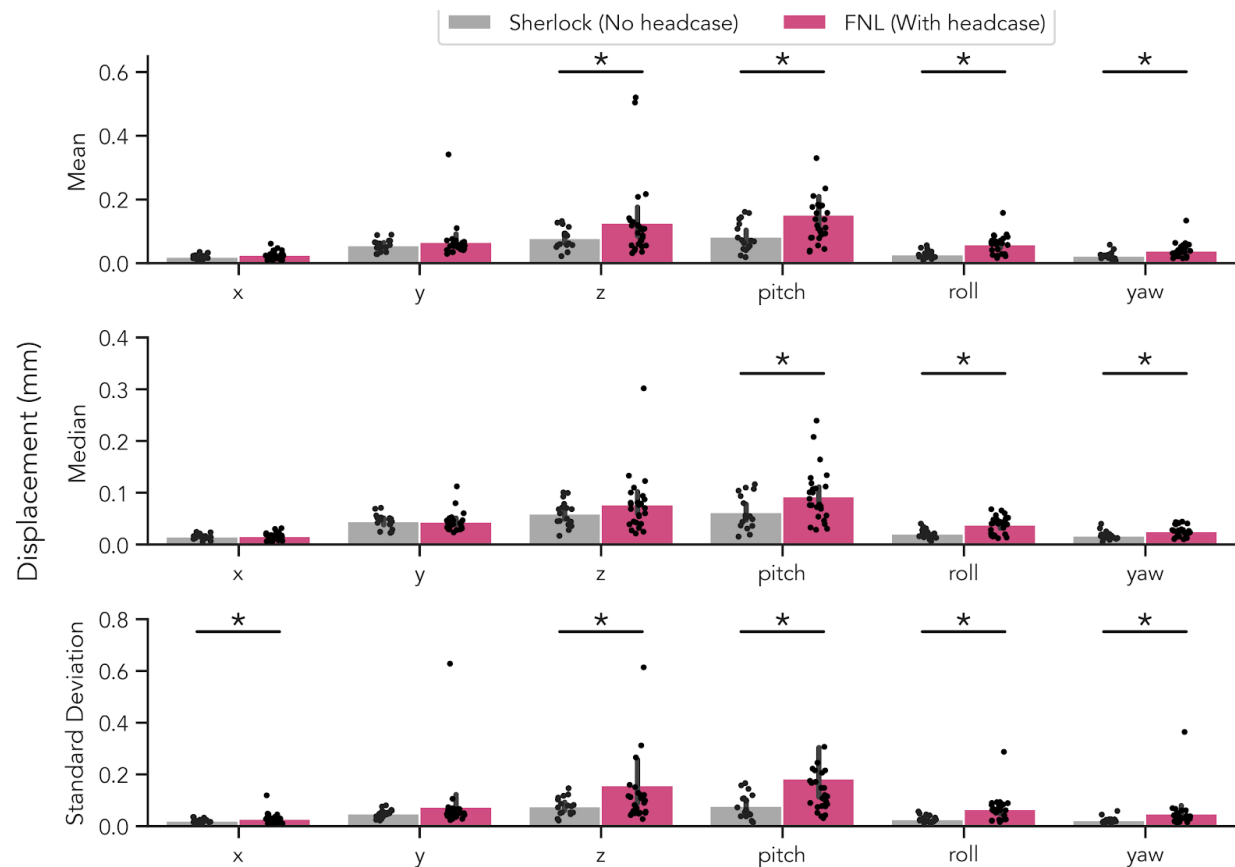
14

508



**Figure 4 | Headcase effect on individual motion parameters while talking**

509  Differences in average displacement for each motion parameter as a function of headcase use while talking (Sherlock
510  and FNL samples). Each plot depicts the absolute value of the backwards-differenced parameter estimate (i.e. the
511  values ultimately summed to computed FD). Top row: average of each participant's mean displacement; middle row:
512  median displacement; bottom row: standard deviation of displacement. In all metrics, headcase participants
513  demonstrated significantly more displacement in all rotation directions (pitch, roll, yaw) with pitch demonstrating the
514  largest differences, likely as a result of bottom-jaw induced head movement. Significant differences were also
515  observed in the z-axis for mean displacement and the x and z-axes for standard deviation of displacement.

516
517

518  **4. Discussion**

519

520  Using three fMRI datasets, we tested the efficacy of custom-molded headcases in reducing
521  head motion during longer naturalistic tasks (viewing a movie or talking aloud). Unlike previous
522  work (Power et al., 2019), which demonstrated an overall reduction in head motion when
523  individuals wore headcases during a brief rsfMRI scan, we found that headcases provided little
524  benefit in reducing head motion. Between group comparisons (with and without headcase) of
525  "task-free" movie-watching indicated no significant reduction in overall head motion indexed by
526  mean and median Framewise Displacement (FD), no reduction in the proportion of high-motion
527  volumes (volumes in which FD exceeded 0.3mm), and no reduction in motion within
528  small-motion volumes only (volumes in which FD was below 0.3mm). These findings are

15

529    unlikely to be driven by sampling idiosyncrasies as participants were recruited from the same
530    population (Dartmouth College community) and scanned using the same equipment with
531    identical acquisition parameters and stimulus. On the contrary, this population as a whole
532    displayed less overall motion relative to a similar population that also underwent a naturalistic
533    movie-viewing paradigm collected at a different site (Princeton University) (Chen et al., 2017).
534    Additionally, overall motion estimates for participants in our samples were approximately similar
535    to participants in the same age range within the sample examined by Power et al (2019).

536

537    Surprisingly, we found that participants who wore headcases produced *more* motion than those
538    who did not during a naturalistic memory recall task, in which they spoke aloud inside the
539    scanner. This seems to have been driven by individual volumes with large movements (FD
540    exceeding 0.3mm), as group comparisons excluding these volumes yielded no significant
541    differences between groups. These larger movements occurred in specific translation directions
542    (i.e., z-axis) and across all rotational axes during talking. These findings suggest that headcases
543    may be inadequate at restricting larger movements and in fact, may paradoxically amplify
544    motion in scenarios that researchers might expect head motion to be worse (talking). With this
545    finding in mind, we speculate that it is possible that by trying to restrict head motion while
546    moving their lower jaws, participants rotate their heads *more* when wearing headcases (relative
547    to foam pillows) increasing overall motion.

548

549    These null differences along with equivalence tests (Lakens et al., 2018) using effect sizes from
550    previous work (Power et al., 2019) highlight the limited efficacy of headcases in reducing head
551    motion. Researchers must balance a tradeoff between the added time, money, and effort of data
552    quality improvement procedures and their expected benefit. For Caseforge headcases
553    specifically, researchers must order a special 3D camera from the company, photograph each
554    participant in a separate session prior to MRI scanning, and await the arrival of each case, and
555    in some cases repeat this procedure if new cases contain defects or fit poorly. In our
556    experience, nearly 20% of participants experienced discomfort when using headcases for
557    extended periods of time precluding their use altogether or requiring adjusted procedures like
558    using the front or back of the case only. This adds additional time and logistical challenges to
559    data collection on top of the existing challenges that collecting MRI data requires. Because
560    headcases are by definition personalized for each individual, they offer no reusability if these
561    same participants are no longer available for future scanning sessions. This may encourage
562    researchers at a given institution to encourage repeated sampling of a small subset of
563    participants for whom headcases exist, decreasing the generalizability of empirical findings
564    (Henrich et al., 2010; Yarkoni, 2019).

565

566    In the datasets analyzed here, each participant without a headcase was situated with foam
567    pillows only (Dataset 3) or both foam pillows and medical tape (Dataset 1). We found this
568    procedure to be adequate and flexible in a non-clinical, non-developmental population without
569    the added burden of acquiring headcases. Interestingly, we found that participants for whom
570    medical tape was used actually produced less overall motion relative to those who were situated
571    with foam pillows alone (*FNL no headcase* vs *Sherlock*). We believe that this was driven by the

16

572    tactile feedback on participants' foreheads provided by the tape rather than any additional
573    movement constraints. This is consistent with recent findings by Kraus et al (2019) who
574    observed less within-run, between-run, and drifting head motion when using medical tape.
575    Specifically, they observed that the efficacy of tactile feedback scaled with the amount of motion
576    observed without tactile feedback, making this procedure particularly appealing when
577    participants are likely to move a lot. Their findings also appear to be independent of behavioral
578    tasks performed in the scanner and are efficacious even in situations when a participant is
579    deliberately asked to move their head (Krause et al., 2019). Another promising alternative is the
580    use of inflatable head pillows (e.g. Pearltec MULTIPAD Positioning System, Newmatic Medical,
581    2020) which may offer a compromise between completely head conforming designs like
582    headcases and more general approaches like foam pillows. However, to our knowledge a
583    systematic investigation using inflatable pillows has not yet been conducted.

584

585    The goal of the present work is to provide an in-depth analysis of the utility of headcases in
586    reducing head motion[3]. Currently, there is only a single published investigation on the efficacy of
587    headcases conducted using a small sample of 14 individuals and a short rsfMRI scan of 4.8
588    minutes (Power et al., 2019). Our work adds to the existing literature to help researchers make
589    a more informed decision about their data collection procedures. We are grateful that
590    companies like Caseforge are working on developing fast, customizable, and accessible
591    pipelines to battle the pernicious issue of head motion for fMRI and hope that these results may
592    lead to improvements in the design and manufacturing processes. Through private
593    communication with Caseforge (Gao, 2019), we were notified that the headcases used in our
594    sample (and presumably those used by Power et al (2019)) were the first generation of their
595    kind and that more improvements to the reliability and comfort of headcases have been made in
596    the second and forthcoming generations. While we are unable to test these claims, they provide
597    promise for reducing the attrition rate we observed and further reducing head motion. Our
598    sample also consisted of a non-clinical, non-development population. Such populations often
599    exhibit increased head motion, a scenario in which headcases may provide more observable
600    benefits  (Vanderwal et al., 2015). The figures presented in Power et al (2019) support this
601    notion as the largest motion reductions occurred for younger participants (7-14 years old). We
602    speculate as to whether training procedures combined with tactile feedback might be similarly
603    successful as participants can be taught to monitor their own movements (Krause et al., 2019).

604

605    In conclusion, we provide data and comparisons that speak to the efficacy of customized
606    headcases under highly demanding data acquisition conditions: long task-free runs of
607    movie-watching and active verbalization. Unlike previous work, we find that customized
608    headcases provide no significant benefits for the reduction of head motion in a non-clinical,
609    non-developmental population. We encourage future researchers to perform additional
610    comparisons using alternative procedures for reducing headmotion (e.g., inflatable pillows,
611    medical tape, headcases, etc). Together these investigations can better help the broader

---

[3] Head cases have also been advertised to ensure that an individual's head is in the same position at the iso-center of the magnetic and head coil across sessions. We were unable to evaluate this in the present work and instead focused on motion mitigation.

612 neuroimaging community adopt the most ideal practices to mitigate motion from contaminating
613 fMRI data (Zaitsev et al., 2015).
614
615 **Open Practices Statement**
616 All code and data required to reproduce the analyses and figures in this manuscript is available
617 on github at https://github.com/cosanlab/headcase and on OSF at https://osf.io/qf6vx. A preprint
618 of this manuscript is available on bioRxiv.
619

633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654

18

655

656 **Table 1. Head case comparisons while *viewing* movies**

| Comparison | Condition | Metric | $M_{Difference}$ | $t$ | $p_{perm}$ |
|---|---|---|---|---|---|
| FNL no-case - FNL with-case | Viewing | $FD_{Mean}$ | -0.016 (-0.063 0.02) | -0.758 | 0.494 |
| FNL no-case - FNL with-case | Viewing | $FD_{Median}$ | 0.007 (-0.009 0.021) | 0.857 | 0.393 |
| FNL no-case - FNL with-case | Viewing | $Spike_{Proportion}$ | -0.015 (-0.049 0.015) | -0.917 | 0.363 |
| FNL no-case - Sherlock | Viewing | $FD_{Mean}$ | -0.008 (-0.032 0.015) | -0.703 | 0.489 |
| FNL no-case - Sherlock | Viewing | $FD_{Median}$ | -0.017 (-0.034 -0.002) | -2.061 | **0.045** |
| FNL no-case - Sherlock | Viewing | $Spike_{Proportion}$ | 0.009 (-0.016 0.033) | 0.723 | 0.466 |
| Sherlock - FNL with-case | Viewing | $FD_{Mean}$ | -0.008 (-0.053 0.031) | -0.349 | 0.776 |
| Sherlock - FNL with-case | Viewing | $FD_{Median}$ | 0.024 (0.007 0.042) | 2.6 | **0.015** |
| Sherlock - FNL with-case | Viewing | $Spike_{Proportion}$ | -0.024 (-0.06 0.009) | -1.375 | 0.185 |

657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673

**Table 2. Head case comparisons while *viewing* movies after excluding high motion TRs**

| Comparison | Condition | Metric | $M_{Difference}$ | $t$ | $p_{perm}$ |
|---|---|---|---|---|---|
| FNL no-case - FNL with-case | Viewing | $FD_{MeanFilterd}$ | 0.008 (-0.005 0.021) | 1.186 | 0.244 |
| FNL no-case - FNL with-case | Viewing | $FD_{MedianFiltered}$ | 0.009 (-0.005 0.021) | 1.33 | 0.192 |
| FNL no-case - Sherlock | Viewing | $FD_{MeanFiltered}$ | -0.011 (-0.025 0.002) | -1.541 | 0.134 |
| FNL no-case - Sherlock | Viewing | $FD_{MedianFiltered}$ | -0.018 (-0.032 -0.004) | -2.368 | **0.022** |
| Sherlock - FNL with-case | Viewing | $FD_{MeanFiltered}$ | 0.019 (0.005 0.034) | 2.43 | **0.022** |
| Sherlock - FNL with-case | Viewing | $FD_{MedianFiltered}$ | 0.026 (0.012 0.042) | 3.319 | **0.002** |

20

701

**Table 3. Head case comparisons while *talking*. Filtered values reflect comparisons after excluding high motion TRs.**

| Comparison | Condition | Metric | $M_{Difference}$ | $t$ | $p_{perm}$ |
|---|---|---|---|---|---|
| Sherlock - FNL with-case | Talking | $FD_{Mean}$ | -0.18 (-0.343 -0.051) | -2.353 | **0.009** |
| Sherlock - FNL with-case | Talking | $FD_{Median}$ | -0.106 (-0.188 -0.029) | -2.534 | **0.013** |
| Sherlock - FNL with-case | Talking | $Spike_{Proportion}$ | -0.171 (-0.306 -0.032) | -2.353 | **0.025** |
| Sherlock - FNL with-case | Talking | $FD_{MeanFiltered}$ | 0.002 (-0.016 0.018) | 0.190 | 0.857 |
| Sherlock - FNL with-case | Talking | $FD_{MedianFiltered}$ | -0.002 (-0.024 0.017) | -0.222 | 0.829 |

704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727

728

729 **Table 4. Head case comparisons while *talking*. Values reflect mean differences and**
730 **bootstrapped 95% confidence intervals for motion parameters without headcase - with**
731 **headcase.**

| Motion Parameter | Metric | Mean$_{Difference}$ | t | p$_{perm}$ |
|---|---|---|---|---|
| X | Mean | -0.005 (-0.011 0.001) | -1.614 | 0.122 |
| Y | Mean | -0.01 (-0.037 0.009) | -0.838 | 0.588 |
| Z | Mean | -0.048 (-0.102 -0.005) | -1.898 | **0.056** |
| Pitch | Mean | -0.069 (-0.132 -0.02) | -2.356 | **0.007** |
| Roll | Mean | -0.031 (-0.045 -0.019) | -4.596 | **< 0.001** |
| Yaw | Mean | -0.016 (-0.028 -0.006) | -2.953 | **0.002** |
| X | Median | -0.001 (-0.005 0.002) | -0.608 | 0.552 |
| Y | Median | 0.001 (-0.009 0.01) | 0.171 | 0.874 |
| Z | Median | -0.017 (-0.043 0.004) | -1.425 | 0.166 |
| Pitch | Median | -0.031 (-0.055 -0.007) | -2.374 | **0.023** |
| Roll | Median | -0.018 (-0.025 -0.011) | -4.674 | **< 0.001** |
| Yaw | Median | -0.008 (-0.014 -0.002) | -2.748 | **0.011** |
| X | SD | -0.009 (-0.019 -0.001) | -1.890 | **0.043** |
| Y | SD | -0.026 (-0.075 0.004) | -1.141 | 0.216 |

22

| | | | | |
|---|---|---|---|---|
| Z | SD | -0.082 (-0.179 -0.01) | -1.838 | **0.040** |
| Pitch | SD | -0.105 (-0.231 -0.026) | -1.896 | **0.004** |
| Roll | SD | -0.038 (-0.061 -0.021) | -3.574 | **0.000** |
| Yaw | SD | -0.027 (-0.057 -0.008) | -2.009 | **0.001** |

732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, *95*(3), 709–721.e5.

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101.

Bettinardi, V., Scardaoni, R., Gilardi, M. C., Rizzo, G., Perani, D., Paulesu, E., Striano, G., Triulzi, F., & Fazio, F. (1991). Head holder for PET, CT, and MR studies. *Journal of Computer Assisted Tomography*, *15*(5), 886–892.

Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Curtis, V. A., Morris, R. G., Williams, S. C. R., Sharma, T., & McGuire, P. K. (1999). Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fMRI. *Human Brain Mapping*, *7*(1), 38–48.

Caseforge Inc. (2020). *Caseforge*. https://caseforge.co/

Chang, L. J., Jolly, E., Cheong, J. H., & Rapuano, K. (2018). Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *bioRxiv*. https://www.biorxiv.org/content/10.1101/487892v1.abstract

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115–125.

Etzel, J. A., & Braver, T. S. (2018). *multibandCFtests* [Data set]. Openneuro. https://doi.org/10.18112/OPENNEURO.DS001544.V1.1.0

Finn, E. S., Glerean, E., Khojandi, A. Y., Nielson, D., Molfese, P. J., Handwerker, D. A., & Bandettini, P. (2019). *Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging*. https://doi.org/10.31234/osf.io/yeu89

24

787 Fitzsimmons, J. R., Scott, J. D., Peterson, D. M., Wolverton, B. L., Webster, C. S., & Lang, P. J.

788     (1997). Integrated RF coil with stabilization for fMRI human cortex. *Magnetic Resonance in*

789     *Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of*

790     *Magnetic Resonance in Medicine*, *38*(1), 15–18.

791 Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996).

792     Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine: Official*

793     *Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic*

794     *Resonance in Medicine*, *35*(3), 346–355.

795 Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., &

796     Ghosh, S. S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data

797     processing framework in python. *Frontiers in Neuroinformatics*, *5*, 13.

798 Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G.,

799     Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li,

800     X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A.

801     (2016). The brain imaging data structure, a format for organizing and describing outputs of

802     neuroimaging experiments. *Scientific Data*, *3*, 160044.

803 Hallquist, M. N., Hwang, K., & Luna, B. (2013). The nuisance of nuisance regression: spectral

804     misspecification in a common approach to resting-state fMRI preprocessing reintroduces

805     noise and obscures functional connectivity. *NeuroImage*, *82*, 208–225.

806 Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I.,

807     Hanke, M., & Ramadge, P. J. (2011). A common, high-dimensional model of the

808     representational space in human ventral temporal cortex. *Neuron*, *72*(2), 404–416.

809 Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based

810     behavioral science. *The Behavioral and Brain Sciences*, *33*(2-3), 111.

25

811 Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust

812 and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2),

813 825–841.

814 Jo, H. J., Gotts, S. J., Reynolds, R. C., Bandettini, P. A., Martin, A., Cox, R. W., & Saad, Z. S.

815 (2013). Effective Preprocessing Procedures Virtually Eliminate Distance-Dependent Motion

816 Artifacts in Resting State FMRI. *Journal of Applied Mathematics*, *2013*.

817 https://doi.org/10.1155/2013/935154

818 Jolly, E. (2018). Pymer4: Connecting R and Python for Linear Mixed Modeling. *Journal of Open*

819 *Source Software*, *3*(31), 862.

820 Jones, E., Oliphant, T., & Peterson, P. (2001--). *{SciPy}: Open source scientific tools for*

821 *{Python}*. http://www.scipy.org

822 Krause, F., Benjamins, C., Eck, J., Lührs, M., van Hoof, R., & Goebel, R. (2019). Active head

823 motion reduction in magnetic resonance imaging using tactile feedback. *Human Brain*

824 *Mapping*, *40*(14), 4026–4037.

825 Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity

826 analysis--connecting the branches of systems neuroscience. *Frontiers in Systems*

827 *Neuroscience*, *2*. https://www.ncbi.nlm.nih.gov/pmc/articles/pmc2605405/

828 Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological

829 Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2),

830 259–269.

831 Meissner, T. W., Walbrin, J., Nordt, M., Koldewyn, K., & Weigelt, S. (2019). Let's take a break:

832 Head motion during fMRI tasks is reduced in children and adults if data acquisition is

833 distributed across sessions or days. *bioRxiv*, 816116.

834 Menon, V., Lim, K. O., Anderson, J. H., Johnson, J., & Pfefferbaum, A. (1997). Design and

26

835    efficacy of a head-coil bite bar for reducing movement-related artifacts during functional

836    MRI scanning. *Behavior Research Methods, Instruments, & Computers: A Journal of the*

837    *Psychonomic Society, Inc*, *29*(4), 589–594.

838  Mowinckel, A. M., Espeseth, T., & Westlye, L. T. (2012). Network-specific effects of age and

839    in-scanner subject motion: a resting-state fMRI study of 238 healthy adults. *NeuroImage*,

840    *63*(3), 1364–1373.

841  Muschelli, J., Nebel, M. B., Caffo, B. S., Barber, A. D., Pekar, J. J., & Mostofsky, S. H. (2014).

842    Reduction of motion-related artifacts in resting state fMRI using aCompCor. *NeuroImage*,

843    *96*, 22–35.

844  Newmatic Medical. (2020). *Pearltec MULTIPAD Positioning System*.

845    https://www.newmaticmedical.com/positioners/p/MULTIPAD-Positioning-System/

846  Optoacoustics. (2020). *FOMRI III+*. http://www.optoacoustics.com/medical/fomri-iii/features

847  Phipson, B., & Smyth, G. K. (2010). Permutation P-values should never be zero: calculating

848    exact P-values when permutations are randomly drawn. *Statistical Applications in Genetics*

849    *and Molecular Biology*, *9*, Article39.

850  Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious

851    but systematic correlations in functional connectivity MRI networks arise from subject

852    motion. In *NeuroImage* (Vol. 59, Issue 3, pp. 2142–2154).

853    https://doi.org/10.1016/j.neuroimage.2011.10.018

854  Power, J. D., Schlaggar, B. L., & Petersen, S. E. (2015). Recent progress and outstanding

855    issues in motion correction in resting state fMRI. *NeuroImage*, *105*, 536–551.

856  Power, J. D., Silver, B. M., Silverman, M. R., Ajodan, E. L., Bos, D. J., & Jones, R. M. (2019).

857    Customized head molds reduce motion during resting state fMRI scans. *NeuroImage*, *189*,

858    141–149.

859   Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E.,

860       Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved

861       framework for confound regression and filtering for control of motion artifact in the

862       preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240–256.

863   Satterthwaite, T. D., Wolf, D. H., Loughead, J., Ruparel, K., Elliott, M. A., Hakonarson, H., Gur,

864       R. C., & Gur, R. E. (2012). Impact of in-scanner head motion on multiple measures of

865       functional connectivity: relevance for studies of neurodevelopment in youth. *NeuroImage*,

866       *60*(1), 623–632.

867   Schuirmann, D. L. (1981). On hypothesis-testing to determine if the mean of a

868       normal-distribution is contained in a known interval. *Biometrics*, *37*, 617–617.

869   Sensimetrics. (2020). *S14 Inset Earphones for fMRI Research*.

870       https://www.sens.com/products/model-s14/

871   Siegel, J. S., Power, J. D., Dubis, J. W., Vogel, A. C., Church, J. A., Schlaggar, B. L., &

872       Petersen, S. E. (2014). Statistical improvements in functional magnetic resonance imaging

873       analyses produced by censoring high-motion data points. *Human Brain Mapping*, *35*(5),

874       1981–1996.

875   Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural

876       systems underlie the production and comprehension of naturalistic narrative speech.

877       *Proceedings of the National Academy of Sciences of the United States of America*,

878       *111*(43), E4687–E4696.

879   Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies

880       successful communication. In *Proceedings of the National Academy of Sciences* (Vol. 107,

881       Issue 32, pp. 14425–14430). https://doi.org/10.1073/pnas.1008662107

882   Thesen, S., Heid, O., Mueller, E., & Schad, L. R. (2000). Prospective acquisition correction for

head motion with image-based tracking for real-time fMRI. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *44*(3), 457–465.

Thulborn, K. R. (1999). Visual feedback to stabilize head position for fMRI. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *41*(5), 1039–1043.

Tyszka, J. M., Kennedy, D. P., Paul, L. K., & Adolphs, R. (2014). Largely typical patterns of resting-state functional connectivity in high-functioning adults with autism. *Cerebral Cortex* , *24*(7), 1894–1905.

Vallat, R. (2018). Pingouin: statistics in Python. *Journal of Open Source Software*, *3*(31), 1026.

Vanderwal, T., Eilbott, J., & Castellanos, F. X. (2019). Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. *Developmental Cognitive Neuroscience*, *36*, 100600.

Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L. C., & Castellanos, F. X. (2015). Inscapes: A movie paradigm to improve compliance in functional magnetic resonance imaging. *NeuroImage*, *122*, 222–232.

Van Dijk, K. R. A., Sabuncu, M. R., & Buckner, R. L. (2012). The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage*, *59*(1), 431–438.

Visconti di Oleggio Castello, M., Dobson, J. E., Sackett, T., Kodiweera, C., Haxby, J. V., Goncalves, M., Ghosh, S., & Halchenko, Y. O. (2018). *ReproNim/reproin: 0.1.1*. https://doi.org/10.5281/zenodo.1207118

Yan, C.-G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R. C., Di Martino, A., Li, Q., Zuo, X.-N., Castellanos, F. X., & Milham, M. P. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *NeuroImage*,

*76*, 183–201.

Yarkoni, T. (2019). *The Generalizability Crisis*. https://doi.org/10.31234/osf.io/jqw35

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., & Hasson, U. (2017). How We Transmit

Memories to Other Brains: Constructing Shared Neural Representations Via

Communication. *Cerebral Cortex* , *27*(10), 4988–5000.

Zaitsev, M., Maclaren, J., & Herbst, M. (2015). Motion artifacts in MRI: A complex problem with

many partial solutions. *Journal of Magnetic Resonance Imaging: JMRI*, *42*(4), 887–901.