Agito ergo sum: in-scanner spatiotemporal motion features reflect anthropometry, behavior and psychiatric function

Authors and affiliations

Thomas A.W. Bolton[1,2], Valeria Kebets[3], Enrico Glerean[4], Daniela Zöller[1,2,5] and Dimitri Van De Ville[1,2]

[1] Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[2] Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland
[3] Singapore Institute for Neurotechnology, National University of Singapore, Singapore
[4] Department of Computer Science, Aalto University, Helsinki, Finland
[5] Developmental Imaging and Psychopathology Laboratory, Office Médico-Pédagogique, Department of Psychiatry, University of Geneva, Geneva, Switzerland

Thomas Bolton 29.4.19 13:58
Comment [1]: Cesar will be added to the authors' list as well, following his feedback (moved from Acknowledgments)!

Corresponding author

Thomas A.W. Bolton
Address: Chemin des Mines 09, 1202 Geneva
Phone: +41 21 69 55 254
Email: thomas.bolton@epfl.ch

Abstract

Freeing resting state functional magnetic resonance imaging (fMRI) data from the deleterious impacts of in-scanner motion has been a hobbyhorse of the neuroimaging community. State-of-the-art guidelines advise to regress out models of nuisance variables, to scrub out excessively corrupted frames as assessed by a composite framewise displacement (FD) score, and to include average FD as a covariate in group-level analyses.

Here, we explored whether the complexity of current motion correction strategies truly fits head motion characteristics. We observed that scrubbed frames relate to very diverse motion profiles, and that the extent to which they are explored relates to a wide array of anthropometric and behavioral features.

On non-scrubbed frames, we found that movers were distinguished by their spatiotemporal profile of motion along the full resting-state session. Strong relationships with behavior were also evident in this setting.

In short, our results call for a shift towards the use of non-aggregated FD measurements at all motion denoising steps in order to avoid sampling biases. They also favor the use of extended sets of regressors modeling frame-to-frame displacement, as opposed to simpler models.

Introduction

Resting-state functional magnetic resonance imaging (RS fMRI) has been a vibrant and flourishing research area. Since its advent (Biswal et al. 1995), the assessment of statistical interdependence between brain regions, or *functional connectivity* (FC), has enabled the determination of large-scale functional brain networks (Damoiseaux et al. 2006, Yeo et al. 2011, Power et al. 2011), and the harvesting of their spatiotemporal properties towards a refined understanding of a constellation of brain disorders (Fox and Greicius 2010).

One of the most remarkable features of RS fMRI is that such analyses are already feasible from as little as 6 minutes of acquisition (Van Dijk et al. 2009). However, the reliance on low amounts of data also requires that the acquired time courses be impeccably cleaned from potential confounding signals. This is even more of a concern as the field starts moving towards even sparser types of analyses, such as dynamic FC (Laumann et al. 2016; see Preti et al. 2018 for a review) or real-time neurofeedback (Watanabe et al. 2017) assessments.

Amongst confounding signal sources, in-scanner motion of the subjects has been a leading cause of investigation. Its deleterious impacts may take many forms, and remain incompletely understood (see Caballero-Gaudes and Reynolds 2017 for a review). The pivotal discovery that even short-lived episodes of motion may greatly bias FC analyses (Power et al. 2012, Van Dijk et al. 2012, Satterthwaite et al. 2012) and lead to erroneous interpretations in clinical or developmental studies (Deen and Pelphrey 2012, Makowski et al. 2019) further fueled the development of robust post-processing strategies to free fMRI time courses from pervasive motion effects.

Through many rigorous and extensive studies (Satterthwaite et al. 2013, Yan et al. 2013, Power et al. 2014, Burgess et al. 2016, Ciric et al. 2017, Parkes et al. 2018), a consensus as for what general steps are essential to a viable RS fMRI denoising pipeline could be reached, although their specificities remain debated. In short, following the linear realignment of functional images (Jenkinson et al. 2002), estimates of motion over time are obtained along three translational directions (X, Y and Z) and three rotational planes (referred to hereafter as α, β and γ). Those estimated motion time courses are then linearly regressed out from the fMRI data, in a matrix of regressors that may, or not, also include their quadratic expansions, their derivatives, and/or their squared derivatives. More parsimonious models lead to a greater amount of retained degrees of freedom in the data, while more exhaustive models may remove signal of interest (Bright and Murphy 2015), but enable to account for biophysically relevant nonlinear motion effects (Friston et al. 1996).

Scrubbing, that is, the exclusion of data points corrupted by excessive instantaneous motion, may be embedded within the regression step (Lemieux et al. 2007) or performed at a later denoising stage (Power et al. 2012). Framewise displacement (FD) is computed as an aggregated measure across all motion parameters[1]. The effectiveness of scrubbing is evident, but the wide variety of possibly resulting fMRI signal changes, in their duration and topology, remains incompletely understood (Power et al. 2014).

Finally, the addition of a covariate for group-level analyses has also been warranted (Ciric et al. 2018). Average FD over time is typically used, and thus, neither spatial nor temporal FD subtleties are accounted for. In addition, this step has been criticized for its risk of biasing some RS fMRI analyses: indeed, head motion correlates with attentional or impulsivity levels (Wylie et al. 2014, Kong et al. 2014), and may thus be a marker of cognitive control abilities (Zeng et al. 2014). It shows clear heritability (Couvy-Duchesne et al. 2014), even if solely non-scrubbed frames are considered (Engelhardt et al. 2017), and shares genetic influences with hyperactivity (Couvy-Duchesne et al. 2016) or body mass index (Hodgson et al. 2017). Correcting fMRI data according to head motion extent may thus prevent from unraveling relevant population differences.

The tight interplay between motion and anthropometric or behavioral factors makes sense when one reflects on the MRI scanner environment: it is expectable that the associated vivid noise and contiguity may be off-putting to the most sensitive individuals sensory-wise. Further, heavyweight subjects may more hardly refrain from moving, and more impulsive people may not manage to remain continuously still. A recent multivariate assessment (Ekhtiari et al. 2019) confirmed those ties, but did not consider spatial or temporal motion specificities (average FD was used as metric of interest).

Since even the most sophisticated motion correction approaches summarized above are still unable to fully remove deleterious motion influences (Yan et al. 2013, Siegel et al. 2017), our first question in the present work was whether current strategies are intrinsically sufficient to capture individual motion characteristics. We observed that it was not the case: at the level of non-scrubbed frames, our cohort of subjects could be separated into four mover subgroups with spatial and temporal motion specificities that are not always/never modeled in subject-level/group-level regression approaches. At the level of scrubbed frames, different flagged time points were associated to diverse motion profiles, which is not captured by existing FD measures.

Our second related question was whether extracted motion features would consist in endophenotypes of anthropometric, behavioral or psychometric aspects. Our results revealed a complex set of yet uncharacterized, overlapping such relationships, demonstrating that the complexity of in-scanner head motion goes beyond present assumptions.

[1] Here, we will be discussing the FD metric suggested by Power et al. (2012), but other alternatives have also been put forward in the past literature (Jenkinson et al. 2002, Van Dijk et al. 2012).

Materials and Methods

Motion Data Acquisition and Preprocessing

We considered a set of 224 healthy subjects from the *Human Connectome Project* (Smith et al. 2013), scanned at rest (eyes open) over four separate 15-minute sessions at a TR of 0.72s. For each session, motion was estimated along the X, Y and Z axes and the α, β and γ planes by a rigid-body transformation using a single-band reference image (*SBRef*) acquired at the start of each session as a reference, and FSL's FLIRT (Jenkinson et al. 2002). It resulted in 6 time courses (one per motion parameter) with 1200 time points each.

Individual motion time courses were differentiated, so that our analyses would focus on instantaneous displacement from time t to time t+1.

Separation Between Non-scrubbed and Scrubbed Frames

Given their link to separate motion correction steps, we considered non-scrubbed and scrubbed frames separately. We used Power's FD definition (Power et al. 2012) at a threshold of 0.3mm. Resorting to a more aggressive (0.2mm) or more lenient (0.4mm) choice to separate frame types did not modify our global findings (see **Supplementary Material** and **Supplementary Figure X** for a more detailed description).

Motion Analysis (Scrubbed Frames)

To analyze scrubbed frames, we thresholded individual motion time courses so that they would highlight one of three states: +1 (excessive positive motion in the considered direction), 0 (tolerable motion), or -1 (excessive negative motion). For each motion parameter, we used $\mu_i \pm \theta\sigma_i$ as thresholds, with μ_i the mean of the motion time course for parameter i, and σ_i its standard deviation. θ was a common threshold across motion parameters, selected as the lowest value that yielded at least one of 6 excessive motion excursions in all scrubbed frames (θ = 0.38). Discussed results remained qualitatively identical using a larger threshold value (θ_2=1.1, leading to 5% of scrubbed frames to remain in meta-state (0 0 0 0 0 0); see **Supplementary Figure X**).

We performed a meta-state analysis of all possible 3^6 motion configurations. Explored states were sorted in descending median occurrence across subjects, and significant occurrence values were determined non-parametrically: for each subject, we generated 19 surrogate datasets in which we circularly shifted individual motion time courses with respect to each other by a random number of samples selected uniformly between 20 and 1180, thus destroying any possible relationship between motion time courses. The lower bound of 20 samples (14.4s) enables to destroy even possibly very long interplays (Power et al. 2014). Meta-state counts were computed across folds, and in each case, the maximum value of the null distribution was used as a significance threshold at α = 0.05. In addition, we also computed meta-state transition probabilities.

To complement the above measures, we also computed two global metrics quantifying the dynamic exploration of the meta-state space: (1) the number of different states visited by each subject over a session, and (2) the total distance traveled in the meta-state space (computed as the sum of L_1 distances between successive meta-states). Those measures were inspired from a past fMRI report exploring the dFC meta-state space (Miller et al. 2016).

Motion Analysis (Non-scrubbed Frames)

For each motion time course, we computed absolute valued instantaneous displacement. Thus, we did not consider the sign of the changes (*e.g.*, moving positively as opposed to negatively in the X direction); this is because initial analyses indicated that positive-valued and negative-valued movements always compensated, to the exception of the X case (two-sided Wilcoxon rank sum test, $p=0.0001$; see **Supplementary Figure X**).

Then, we averaged motion values within a) each motion type (X, Y, Z, α, β and γ) and b) each of 6 even duration time intervals along the scanning sessions (2.4 min each). This resulted in a total of 36 conditions. We chose 6 temporal sub-bins to give equal weight to spatial and temporal domain information in our decomposition of the data. Eventually, the data was z-scored across subjects for each condition, so that positive values highlight strong movers (at a given time and for a given motion parameter) with respect to the mean, and *vice versa*. It also follows that an equal weight is given to each condition.

To separate all subjects into different subgroups of movers, we performed spectral clustering (Von Luxburg 2007) on an N-nearest neighbor graph ($N = 10$, cosine distance used as distance measure; note that the results remained almost identical with $N = 5$ or $N = 15$, as can be seen in **Supplementary Figure X**). We used the first three eigenvectors with non-zero eigenvalue for the clustering, in order to also be able to represent spatiotemporal motion data graphically (see **Figure 2A/B**).

The optimal number of clusters into which to subdivide the data was assessed through consensus clustering (Monti et al. 2003), a subsampling-based assessment of robustness. In more details, the clustering process was repeatedly run (100 times) over 80% of the data points (*i.e.*, 179 subjects), for cluster numbers k ranging from 2 to 17. For each k, a consensus matrix summarizing how frequently two data points would be clustered together was derived. Since the goal in a good clustering scheme is to either always cluster two data points together, or to never do so, the goal is to find a k for which the proportion of ambiguously clustered pairs (Şenbabaoğlu et al. 2014), linked to intermediate consensus values, is the lowest. As can be seen in **Supplementary Figure X**, $k=4$ stood as a clear optimum.

To evaluate whether there was any significant effect of scanning duration, motion parameter or mover subtype, or any interaction between these factors, we conducted a three-way ANOVA (factor 1: scanning duration, factor 2: motion parameter, factor 3: mover subtype) and assessed significance by comparing the obtained F-values with a null distribution generated non-parametrically over 10'000 folds. In addition, we also individually plotted scanning duration or motion parameter against cluster assignments (**Figure 2C**), averaging over all entries from the other factor (*e.g.*, the bar labeled 'X' denotes the average of motion along the X direction from t_1 to t_6).

Behavioral Data Acquisition

For each subject, a battery of behavioral and demographic scores was also quantified. A list of all the investigated scores in the present study can be found in **Supplementary Table 1**. They were subdivided into several key sub-domains, largely following the original classification found in the HCP Data Dictionary[2]:

1. Demographic parameters, including race, ethnicity, employment status, income or education level
2. Physical health, such as weight, height, body mass index (weight/height^2), blood pressure, hormonal levels
3. Alertness levels, assessed in terms of cognitive status (MMSE; Folstein et al. 1983) and sleep quality (PSQI; Buysse et al. 1989)
4. Cognitive abilities (in terms of accuracy, response time or errors) across various tasks spanning different cognitive domains (see Barch et al. 2013 for details)
5. Emotional level in terms of anger, fear, stress or life satisfaction (assessed through the NIH toolbox; Gershon et al. 2010)
6. Motor abilities, including endurance, gait speed, dexterity and strength measurements
7. Sensory levels, quantified in terms of responses to noise, odor, pain, taste, or contrast
8. Personality traits, as assessed by the NEOFAC questionnaire (McCrae et al. 2004)
9. Psychiatric and life function (Achenbach 2009)
10. Substance use, that is, intake of alcohol, tobacco or drugs (partly from the SSAGA questionnaire; Buchholz et al. 1994).

For some scores, several entries were not acquired in a sub-fraction of subjects (mean: 1.52%, median: 0.89%, maximum: 21.43%). This was taken into account in behavioral data processing (see **Section Behavioral Data Processing**) so that it would exert a minimal effect on the described findings.

Of note is that we did not include some types of scores into our analysis. This included:

1. Gender and age
2. Family relationships between subjects and twin status
3. Psychiatric history of the mother or father
4. Scores reflective of the menstrual cycle in female subjects.

In several cases, *Age adjusted* and *Unadjusted* scores are provided. Both were kept, and aggregated in further processing steps (see **Section Behavioral Data Processing**).

---

[2] https://wiki.humanconnectome.org/display/PublicData/HCP+Data+Dictionary+Public-+Updated+for+the+1200+Subject+Release

Behavioral Data Processing

Because some domains included many more scores than others, and in order to conduct a balanced analysis, we converted the scores of each domain into only one value through Probabilistic Principal Component Analysis (PPCA; Bishop 1999), which enabled, at the same time, to fill in the few missing entries in each case. Some scores were not retained because they appeared irrelevant to us (labeled '2' in **Supplementary Table 1**; *e.g.*, *has blood been sampled?*), and others because they were considered too specific (that is, would induce overfitting), or overlapped with others (labeled '3'; examples would be the individual DDISC scores, which are already summarized in AUC measurements).

Eventually, we only retained the domain scores that were sufficiently accurate, according to the criteria proposed by Smith et al. (2015); excluded scores at this stage are labeled '4'. The final matrix of behavioral information used for the analyses had size 224 x 46 (subjects x domains). Detailed information about what fraction of each domain score is accounted for by a given original variable are provided in the **Supplementary Material** (**Supplementary Figure X**).

Univariate link between motion subgroups and anthropometry/behavior

To determine whether some anthropometric/behavioral domains would differ across mover subgroups, we performed a univariate assessment where, for each of the 46 assessed domains, we computed a score indicative of cluster-to-cluster distinction. Formally, following Gu et al. (2012):

$$F(x^i) = \frac{\sum_{k=1}^{K} n_k (\mu_k^i - \mu^i)^2}{\sum_{k=1}^{K} n_k (\sigma_k^i)^2},$$

where x_i is the vector of the i^{th} domain scores across subjects, μ^i is its average regardless of group classification, μ^i_k is its average within group k, and σ^i_k is the standard deviation within group k. A large score value indicates that the assessed behavioral domain shows distinct values between clusters.

To non-parametrically extract the significant scores, we recomputed each of them 1'000 times after shuffling subject motion entries, and considered for interpretation the domains for which the score exceeded the 95^{th} null distribution percentile following Bonferroni correction for 46 tests.

Multivariate links between motion features and anthropometry/behavior

We performed a multivariate assessment on the extracted motion features from non-scrubbed or from scrubbed frames. We considered the matrix of behavioral scores (size 224 x 46), and the matrix of motion scores (size 224 x 36 for the non-scrubbed case, and 224 x 2 for the scrubbed case). We used Partial Least Squares (Krishnan et al. 2011) to derive components (that is, linear combinations of motion or behavioral scores) showing maximal covariance.

To assess significance of the components, we compared their singular values to a null distribution constructed from 1'000 shuffled datasets, following Zöller et al. (2017). We focused our interpretation on the extracted components significant at α=0.02. Additional component examples are provided in **Supplementary Figure X**.

To determine the significance of individual latent scores, we performed bootstrapping with 80% of the data. At each bootstrapping fold, singular matrices were linearly aligned to the ones from the full decomposition with a Procrustes transform (Gower 1975). Let R_1 the rotation matrix used to align the spatiotemporal motion data, and R_2 the rotation matrix used for behavioral parameters; the final rotation applied was given by $(R_1+R_2)/2$.

To interpret the extracted components in the case of non-scrubbed frames, we converted the 36-element vector resulting from the PLS process into a 6-element space and a 6-element time representation, by averaging across all time points or across all spatial directions, respectively. Z-scores for each entry were computed on those summarizing metrics.

Results

Given the fact that scrubbed frames and non-scrubbed frames are typically separated in preprocessing strategies, we analyzed them independently. Power's framewise displacement metric (FD_{Power}; Power et al. 2012) was used to tag the time points to scrub, and the results below are presented for a threshold value of 0.3mm.

Scrubbed frames arise from diverse types of motion excursions

First, we considered in which directions subjects would move the most during scrubbed frames. **Figure 1A** depicts the top 50 meta-states (ordered in descending order of median expression across subjects) present in our analyzed population. Unsurprisingly, the *no motion* meta-state (0 0 0 0 0 0) was the most prominent. Some combinations of motion parameters were particularly prominent: excursions along the X and the Z directions could be seen in the large majority of top meta-states, and this was accompanied by excursions along the α rotational plane, with opposite sign. Other parameters revealed no obvious relationship from this introductory assessment.

Next, we assessed the recruitment of positive and negative motion excursions along meta-states of decreasing median occurrence (**Figure 1C**). The dashed diagonal line reflects the expected cumulative distribution if a given motion parameter shows excursions uniformly distributed, regardless of meta-state importance. We observed deviations from this case regarding X and Z motion increases, where the most occurring meta-states (up to around 20% of the total pool; first brown vertical bar) showed a larger recruitment than expected (see the positive hump of associated curves). Such relationships were also captured in the case of negative excursions; however, prominent recruitment of meta-states with negative Z direction motion ended earlier, while the opposite was true for the X direction (second brown vertical bar), evidencing an intriguing asymmetry. The presence of negative changes along the β and γ planes was also seen alongside the latter X excursions.

Interestingly, turning to the percentage of subjects showing significant occurrences of the meta-states (**Figure 1B**), three sub-pools could be distinguished: the first involved the 150 most prominent meta-states, with elevated occurrences across subjects. The second showed still notable recruitment of meta-states 150 to around 350, albeit at a lower level. The third included the rest of poorly expressed meta-states. Those transitions occurred at similar meta-state fractions as compared to the above cumulative results.

Mean transitions across meta-states showed a clearly non-random structure (**Figure 1D**). Of note is that apart from transitions from meta-state (0 0 0 0 0 0) to itself, *median* transition probabilities across subjects were all null, highlighting the large heterogeneity in motion dynamics across subjects. Focusing on the top 50 meta-states (**Figure 1D**, lower matrix), meta-states 1, 10, 13, 35-38 and 45 stood amongst the top start and end states. Meta-state 1 is the *no motion* condition, meta-states 10 and 13 solely include X motion, meta-states 35-38 and 45 only include translational motion changes. The top transitions were between meta-states 10 and 24 (from negative X to negative α), 10 and 45 (from negative X to positive α), 13 and 13, and 13 and 36 (positive X to negative Y).

Regarding global meta-state dynamics (**Figure 1E**), subjects showed a wide array of traveled distance (median 6214, minimum 2332, maximum 8336) and visited states (median 368, minimum 194, maximum 498).

Spatiotemporal diversity of motion in non-scrubbed frames

Average motion across six even-duration session bins, and the 6 motion parameters, was quantified on non-scrubbed frames (see **Materials and Methods**). This spatiotemporal motion profile characterization revealed the existence of four separate subgroups of movers (**Figure 2B**): in the first one (n_1=70, red patches), subjects showed low motion across all time and motion dimensions (negative z-score values in **Figure 2C**). In the second (n_2=51, dark blue patches), subjects moved little, and less following the first sixth of the session, with particularly strong motion along the α and β rotational components. The third group (n_3 = 67, orange patches) showed very strong motion spatiotemporally, and the fourth one (n4 = 36, cyan patches) showed particularly strong motion in the γ rotational plane, which slightly attenuated after the first sixth of the session.

Statistical analysis confirmed the above assessments: on top of a significant effect of group (F = 1414.41), there was a significant time x group interaction (F = 3.11), and *post-hoc* assessment revealed that while groups 1, 2 and 4 showed a decrease in motion over time (β1=-0.007 [-0.01,-0.003], p=0.0026, b2=-0.0194 [-0.027,-0.012], p=6.31 10^{-6}, b4 = -0.025 [-0.031,-0.019], , p=5.6 10^{-10}, respectively), group 3 exhibited an increase (b3 = 0.0354 [0.014,0.057], p=0.0013). Thus, different mover subgroups displayed varying temporal changes in their extent of motion.

In terms of spatial properties, there was a significant effect of space (F=5.92), as well as a significant space x group interaction (F = 83.88). Exhaustive results from a *post-hoc* assessment are displayed in **Table 1**. They show that subjects in group 4 moved the most in the gamma plane (hence their blue shade in **Figure 2A**), while subjects from group 2 moved the least there (hence their red and green tones). Group 1 featured the lowest movers in Y, Z, alpha and β, while in group 3, subjects moved most in X, Y, Z, α and β (thus, they appear in white in **Figure 2A**). Overall, each group could thus be clearly distinguished on the basis of spatial motion properties.

Univariate links of mover subgroups to anthropometry and behavior

Next, we related the spatiotemporal motion characteristics of the subjects (as summarized by their mover group assignment) to their anthropometric, behavioral and psychometric features (see **Materials and Methods** for details). **BMI (good)** (which mostly reflects the influence of height), **BMI (bad)** (denoting the impact of weight), **Blood pressure** and **Cognitive flexibility** components were significantly different across mover subtypes following Bonferroni correction (**Figure 3**).

Multivariate assessment of motion/behavior relationships

Finally, we attempted to extract significant relationship(s) between our motion characteristics and anthropomorphic/behavioral/psychometric features (see **Materials and Methods** for methodological details).

Regarding non-scrubbed motion characteristics (**Figure 4**), there were three significant components (p=0, p=0.017, p=0.008). Component 1 characterized motion regardless of space or time (apart from α, all values had an absolute z-score larger than 3): subjects showing larger global motion values also showed worse endurance, larger sensitiveness to sound noise, worse working memory performance, worse spatial orientation (both in terms of larger response time and lower accuracy), lower cognitive flexibility, worse fluid intelligence, and worse body mass index (decreased height, enhanced weight). In addition, they were more anxious, showed more thought problems, aggressiveness, inattention and antisocial behaviors.

Component 2 largely corresponded to movement in the γ plane. Stronger movers in that plane also showed sleep problems, but performed better in cognitive flexibility and spatial orientation tasks. Their response time was larger when tested for sustained attention, and they showed a constellation of elevated psychometric scores including anxiety, withdrawal, somatic problems, attentional problems, aggressiveness, hyper-responsiveness and rule breaking behaviors. They also showed stronger alcohol consumption.

Component 3 contrasted translational (mostly X) and rotational (mostly α) motion. Stronger translational movers (*i.e.*, also lower rotational movers) showed larger weight and height, larger blood pressure, enhanced cognitive flexibility abilities, significant self-regulation abilities, and greater response time upon a sustained attention task. They were also better at discriminating emotions, but worse at contrast sensitivity. Psychometrically, they were more intrusive and showed less externalizing, withdrawal and thought problems.

Regarding the results for scrubbed frames (**Figure 5**), both components were significant (p=0.0011 and p=0.0099). The first one denoted a relationship between lowered explored distance and number of visited meta-states, and enhanced "negative characteristics" reminiscent of Component 1 described above (compare **Figure 4A** and **Figure 5A**): in particular, subjects expressing this component had greater weight, lowered height, worsened cognitive abilities, and globally larger psychmetric scores.

As for the second component (**Figure 5B**), more strongly expressing subjects traveled less in the state space, but explored a broader array of distinct meta-states. By far, the most prominent anthropomorphic feature was a larger height (resulting in a lower body mass index). Interestingly, larger values were also seen specifically for rule breaking and hyper-responsiveness psychometric scores, alcohol consumption, as well as significantly worsened self-regulation.

Discussion

We observed that head motion in the MRI scanner during RS acquisitions, one of the leading topic of fMRI data preprocessing efforts, exhibits a spatiotemporal complexity that goes beyond what is accounted for by commonly used denoising strategies.

At the level of scrubbed frames, many different combinations of motion excursions along all 6 available degrees of freedom were consistently present across subjects. The most prominent directions were X and Z, which match lateral and nodding motions, known to be particularly prominent during scanning. Interestingly, though, there was an asymmetry in that positive changes in X were strongly represented in only the most occurring meta-states (together with positive Z changes), while negative changes in X spanned a wider array of meta-states and appeared to occur alongside β and γ rotational displacement. This raises the possibility of a dual involvement of translational displacement along X: one component, linked to the top meta-states only, would relate to typical in-scanner motion, while the other would solely involve negative translational displacement and may be related to other features. We hypothesized a possible explanatory role of handedness scores, but no significant correlation was found with the bias in positive *versus* negative X recruitment.

Thomas Bolton 29.4.19 14:31

Comment [2]: I need to confirm that X and Z truly reflect those two motion types; also, any idea of a reference where they are acknowledged as particularly prominent (obvious from non-normalized motion values).

On top of the variety in simultaneous changes seen across motion directions, there was also a very structured temporal dynamics, in which transitions did not only involve direct returns to the baseline state. This has several important implications: first, it may partly explain the wide spatiotemporal diversity in observed fMRI signal changes following micro-movements (Power et al. 2014), although other factors are likely to contribute as well, such as the exact timing of motion along the acquisition of successive imaging planes. In order to further investigate the contributions of those different factors, future analyses should focus on motion data obtained at higher temporal resolution, through an external apparatus rather than indirectly from fMRI recordings themselves.

Second, the presence of temporal relationships between successive scrubbed frames also conceptually challenges the notion of scrubbing, which is a frame-level excision strategy and thus, does not incorporate any information about temporality. As RS acquisitions are achieved at increasingly lower TR, this conceptual mismatch is expected to become more and more apparent.

Although this does not argue against the efficiency of scrubbing *per se* (the signals will be cleaned as long as all corrupted frames are removed), it may prevent the implementation of more tailored strategies for motion correction, in which excised fMRI volumes could perhaps be interpolated meaningfully and thus, prevent the loss of degrees of freedom (distinct across subjects) that presently undermines scrubbing. To date, such approaches remain rare (Patel et al. 2014, Yang et al. 2019). An interesting alternative direction to follow may be the application of probabilistic approaches that directly model temporality, such as through hidden Markov models (HMMs), which are already used at the level of RS fMRI analyses (Vidaurre et al. 2017, Bolton et al. 2018).

Time points retained for analysis (*i.e.*, not scrubbed out) showed clear spatiotemporal properties, with changing spatial motion along the course of scanning. Our separation of subjects into four distinct subgroups of movers confirms an expectable, but never explicitly verified fact: that within a population of healthy subjects within a narrow age range, different individuals will show distinct, characteristic and non-random head motion profiles.

More specifically, two mover subgroups were opposite extremes: one moved very little as compared to the average across time and space, while the other consistently displayed very large displacements. Rotational motion was the distinguishing feature of the last two groups, respectively in the α/β (group 2) and the γ (group 4) planes. All in all, the presence of strongly differing spatial motion profiles across subjects confirms the importance of subject-level motion correction through regression. Further, since our work focused on instantaneous motion (t to t+1 changes), our results are an additional argument in favor of more complete regression models, at least to the point of incorporating motion time courses and their shifted counterparts.

Albeit less evident than the spatial component, a temporal change along the scanning session (and its interaction with the space factor) was also statistically significant: this was in largest part due to a changed motion extent after the first sixth of the session. This need for a few minutes before setting into a *motion steady state* suggests that to avoid one possible source of bias, if affordable, future fMRI analyses may be performed on time courses trimmed by a few minutes. Alternatively, it could also be envisaged to explicitly model this early transition as one or several motion regressors.

While subjects from groups 1, 2 and 4 moved less after the first sixth of the recording session, high movers from group 3 moved more. Since the latter stood out in terms of body mass index and blood pressure, a reasonable assumption is that this enhanced motion extent reflects an inability to cope with anthropometric properties over an extended time. As for low movers, they showed significantly larger **BMI (good)** (reflecting the positive influence of height) and **Cognitive flexibility** scores, implying that their reduction in motion (and overall globally low motion across space and time) may be supported by anthropometric factors, as well as by a better ability to cope with the constraints of scanning (for example, perhaps by better adjusting to noise changes over scanning).

Intriguingly, although cognitive flexibility differed between the low movers and the high movers and α/β movers, there was no difference with the γ movers, although those moved more on average than α/β movers. This raises the concern that important aspects regarding head motion may have been missed by our univariate analysis. One possibility could be that our statistical correction was too stringent, since several of the investigated domain scores are actually correlated. Along this line of reasoning, inattention (previously related to head motion as quantified by averaged FD over time; Wylie et al. 2014, Kong et al.

Thomas Bolton 29.4.19 14:34
Comment [3]: BE SURE THAT IT WAS NEVER EXPLICITLY VERIFIED BEFORE: any reference anywhere?

Thomas Bolton 25.4.19 03:03
Comment [4]: I need to figure out what this may reflect exactly, but I am clueless for now as for how to properly discuss it... Any help is very welcome!

Thomas Bolton 29.4.19 14:38
Comment [5]: I should add that explicitly as a Supplementary Figure matrix.

2014) was close to significance and may be one of those additional contributing factors.

A second possibility is inherent to the approach: univariate analysis does not enable to reveal possibly interacting factors which would not be effective on their own, but become meaningful when combined together. To explore this possibility, we complemented our initial assessment by performing a PLS analysis. The first derived component showed remarkable similarities with the aforementioned findings: it characterized motion in all spatial directions, at all time points, with an increase following the first session sixth. Amongst important anthropometric and behavioral domains, body mass index, blood pressure and cognitive flexibility were amongst the dominating factors. Lowered endurance was also revealed in larger movers, on top of additional psychometric scores for aggressiveness, inattention and antisocial behaviors. Larger movers also performed more poorly in spatial orientation, fluid intelligence and working memory tasks. Overall, this general pattern is highly reminiscent of a *positive-negative mode of population covariation* previously described by Smith and colleagues (2015), and put forward as relating behavior, demographics and FC. Our results raise the possibility that this mode, at least in part, reflected differences in motion across the considered subjects (note that the authors considered the same HCP dataset as we did).

Particularly interesting is the observation that the first component found from our analysis of non-scrubbed frames substantially overlapped with that extracted from the analysis of meta-state space exploration on the basis of scrubbed frames. There was a significant positive correlation between mean FD and latent scores of that latter component, indicating that it indeed reflects the behavioral, anthropometric and psychometric properties of large movers. This enables to tie non-scrubbed and scrubbed frames together as partly encompassing similar information about subjects. To further demonstrate this feat, we plotted both meta-parameters in a two-dimensional representation where color coding reflects the group labels extracted from non-scrubbed analyses: the segregation of large mover data points from other groups is evident along the first axis (number of explored meta-states).

Thomas Bolton 29.4.19 14:45
Comment [6]: I could perhaps quantify the agreement between components?

On top of those commonalities, a major asset of PLS is the ability to disentangle linearly overlapping motion/behavior relationships, and as such, to also reveal subtler relationships more specific to one or the other subtype of frames. Accordingly, in our analyses of non-scrubbed frames, on top of component 1, two other more subtle components were exposed: component 2 specifically showcased γ motion, clarifying its link to other motion factors: those subjects that move more along γ also move less along other directions (as indicated by the negative signs in **Figure 4B**). Those *γ movers* were not related to any anthropometric quantity, but showed sleep disturbance, high cognitive flexibility and spatial orientation performance, but also larger sustained attention response times. This was accompanied by a wide scope of elevated psychometric scores, including anxiety, somatic problems, aggressiveness, intrusiveness, ADHD and hyper-responsiveness, as well as by alcohol consumption. We conjecture that this component reflects a *hyper* tendency, featuring agitated subjects who will be disinhibited, efficient at *on the moment* tasks, but less capable when it comes to temporally sustained abilities.

I need something clever about component 3...

As for scrubbed frames, the second component clearly highlighted lower movers, as confirmed by a significant negative correlation between mean FD and latent scores. Hence, subtler effects normally overshadowed by large motions could be revealed, demonstrating that a larger repertoire of meta-states is visited by those subjects that are both taller (as reflected by a large **BMI (good)** z-score) and exhibit a larger tendency towards rule breaking, greater hyper-responsiveness, and worsened self-regulatory abilities. Thus, those subjects likely do not obey the *stay still* guideline as thoroughly, and are particularly *at risk* to yield larger motion effects due to their greater height. They travel a lower overall distance because the meta-states that they visit are similar (that is, differ only across one or two motion parameters), yet they explore a broader array of such configurations.

In 1644, René Descartes, in quest for a primal principle at the root of all knowledge, formulated his notorious *cogito ergo sum* (I think, hence I am)[3]. 375 years later, we wish to summarize our findings by reformulating his words: ***agito ergo sum*** (I move, hence I am). By this, we mean that the defining aspects of someone (one's bodily features, abilities to interact with the world and way to respond to the environment around) are reflected, in subtle and various ways, in how he or she moves during scanning.

This has strong implications regarding RS fMRI studies: indeed, the observation that a broad array of behavioral and psychometric characteristics relate to motion implies that the scope of studies reporting possibly biased findings with regard to clinical or cognitive group-level comparisons is perhaps much wider than envisaged so far. On top of already questioned results of fluid intelligence (Finn et al. 2015; see **Figure 6** of Siegel et al. 2017), former reports focusing on sustained attention (Rosenberg et al. 2016) or extraversion (Hsu et al. 2018) may also need to be reconsidered. At the end of the day, we are faced with a sort of *chicken-and-egg problem*: when resorting to regression strategies, do we get rid of deleterious motion effects, or do we, perhaps, also remove neurally meaningful signal?

In addition, the presence of spatially distinct pools of movers (and their relationships to behavior/psychometrics) may also cast some doubts towards the accuracy of most motion denoising assessment approaches, which exclusively rely on averaged FD over time. A separate assessment across motion parameters (or more elaborate approaches involving specific combinations) appears to be necessary to better understand which motion impacts are removed, and which subsist in the data. Given our findings, perhaps the motion leftovers remaining after even the most optimal preprocessing approaches

---

[3] The first mention of that particular formulation indeed dates back from the *Principia philosophiae,* published in 1644.

reflect the subtler components unraveled here (average FD would, indeed, be expected to mostly be sensitive to the global motion factor, seen as component 1), that is, overshadowed motion/behavior relationships generally masked beneath larger motion contributions (as seen with component 2 of our scrubbed frames analyses).

Some important limitations of our work should be highlighted. First, we explored regression and scrubbing, but did not consider other motion correction alternatives; they include original twists on traditional regression designs (Patriat et al. 2015, Patriat et al. 2017), more sophisticated variants over scrubbing (Patel et al. 2015, Yang et al. 2019), and methods relying on an ICA decomposition of the data (Salhimi-Korshidi et al. 2014, Pruim et al. 2015).

Second, we have not yet pushed our exploration to the level of fMRI time courses themselves, but focused on motion estimates only. Our aim, with this report, was not to design a new efficient motion correction strategy, but to dig into the complexity of motion *per se*, and by this mean, put forward possible caveats and improvements of existing approaches. Our codes and results are fully available at https://c4science.ch/source/MOT_ANA.git, and we encourage the interested researchers to extend our current investigations the fMRI signals themselves.

Third, we solely considered motion, although many more factors are known to corrupt the fMRI signal (Biancardi et al. 2009, Birn 2012, Liu 2016). Particularly relevant to the present study is the recent work of Power et al. (2019), who showed that motion time courses from the HCP dataset contain an array of respiratory contributions. Given the impact of blood pressure on some of our components, it seems likely that cardiac or respiratory effects indeed contribute to head motion variability.

Future motion correction strategies shall improve over current ones in several ways: first, through more elaborate acquisition schemes, such as with multi-echo sequences (Power et al. 2018); second, through the exploration of other complementary denoising strategies, such as with fMRI simulators (Drobjnak et al. 2006) or prospective correction (Zaitsev et al. 2017); third, and perhaps most importantly, through an efficient cross-talk across those strategies. For example, it was recently shown that the use of customized head molds reduces motion during scanning on young subjects (Power et al. 2019); this could be pushed further by orienting the design in subject-specific manner, using motion characteristics such as the ones described here.

Conclusion

In conclusion, we have shown that motion in the fMRI scanner during resting-state acquisition is exquisitely complex, in its spatial as well as its temporal nature. This was the case regardless of whether we considered typically excised time points, or the ones kept for subsequent analyses. We revealed how this spatiotemporal complexity of motion tightly relates to the anthropometric properties, behavioral specificities, and psychometric features of the subjects, and thus, wish for future clinical or cognitive fMRI studies to account for motion-related caveats in more elaborate manners than currently done.

Acknowledgments

The authors would like to thank César Caballero-Gaudes for his valuable insight regarding the presented findings.

Authors' contributions

TB designed the study, ran the analyses and wrote the manuscript. VK contributed to the cognitive interpretation of the findings. EG suggested some of the included analyses. DZ provided the PLS software used in the work, as well as methodological insights. DVDV supervised the work. All authors thoroughly reread the manuscript.

References

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic resonance in medicine*, *34*(4), 537-541.

Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the national academy of sciences*, *103*(37), 13848-13853.

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... & Fischl, B. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, *106*(3), 1125-1165.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., ... & Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, *72*(4), 665-678.

Fox, M. D., & Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Frontiers in systems neuroscience*, *4*, 19.

Van Dijk, K. R., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., & Buckner, R. L. (2009). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *Journal of neurophysiology*, *103*(1), 297-321.

Laumann, T. O., Snyder, A. Z., Mitra, A., Gordon, E. M., Gratton, C., Adeyemo, B., ... & McCarthy, J. E. (2016). On the stability of BOLD fMRI correlations. *Cerebral cortex*, *27*(10), 4719-4732.

Preti, M. G., Bolton, T. A., & Van De Ville, D. (2017). The dynamic functional connectome: State-of-the-art and perspectives. *Neuroimage*, *160*, 41-54.

Watanabe, T., Sasaki, Y., Shibata, K., & Kawato, M. (2017). Advances in fMRI real-time neurofeedback. *Trends in cognitive sciences*, *21*(12), 997-1010.

Caballero-Gaudes, César, and Richard C. Reynolds. "Methods for cleaning the BOLD fMRI signal." *Neuroimage* 154 (2017): 128-149.

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

Power, Jonathan D., et al. "Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion." *Neuroimage* 59.3 (2012): 2142-2154.

Satterthwaite, T. D., Wolf, D. H., Loughead, J., Ruparel, K., Elliott, M. A., Hakonarson, H., ... & Gur, R. E. (2012). Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. *Neuroimage*, *60*(1), 623-632.

Deen, B., & Pelphrey, K. (2012). Perspective: brain scans need a rethink. *Nature*, *491*(7422), S20.

Makowski, C., Lepage, M., & Evans, A. C. (2019). Head motion: the dirty little secret of neuroimaging in psychiatry. *Journal of psychiatry & neuroscience: JPN*, *44*(1), 62.

Yan, C. G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R. C., Di Martino, A., ... & Milham, M. P. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *Neuroimage*, *76*, 183-201.

Thomas Bolton 25.4.19 04:32
Comment [7]: I need to make it proper style and add a few that I did not fetch from Scholar yet...

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., ... & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage*, *64*, 240-256.

Power, Jonathan D., et al. "Methods to detect, characterize, and remove motion artifact in resting state fMRI." *Neuroimage* 84 (2014): 320-341.

Burgess, G. C., Kandala, S., Nolan, D., Laumann, T. O., Power, J. D., Adeyemo, B., ... & Barch, D. M. (2016). Evaluation of denoising strategies to address motion-correlated artifacts in resting-state functional magnetic resonance imaging data from the human connectome project. *Brain connectivity*, *6*(9), 669-680.

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., ... & Gur, R. C. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage*, *154*, 174-187.

Parkes, L., Fulcher, B., Yücel, M., & Fornito, A. (2018). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *Neuroimage*, *171*, 415-436.

Ciric, R., Rosen, A. F., Erus, G., Cieslak, M., Adebimpe, A., Cook, P. A., ... & Satterthwaite, T. D. (2018). Mitigating head motion artifact in functional connectivity MRI. *Nature protocols*, *13*(12), 2801.

Jenkinson, Mark, et al. "Improved optimization for the robust and accurate linear registration and motion correction of brain images." *Neuroimage* 17.2 (2002): 825-841.

Bright, M. G., & Murphy, K. (2015). Is fMRI "noise" really noise? Resting state nuisance regressors remove variance with network structure. *NeuroImage*, *114*, 158-169.

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S., & Turner, R. (1996). Movement-related effects in fMRI time-series. *Magnetic resonance in medicine*, *35*(3), 346-355.

Lemieux, L., Salek-Haddadi, A., Lund, T. E., Laufs, H., & Carmichael, D. (2007). Modelling large motion events in fMRI studies of patients with epilepsy. *Magnetic resonance imaging*, *25*(6), 894-901.

Siegel, J. S., Power, J. D., Dubis, J. W., Vogel, A. C., Church, J. A., Schlaggar, B. L., & Petersen, S. E. (2014). Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Human brain mapping*, *35*(5), 1981-1996.

Miller, R. L., Yaesoubi, M., Turner, J. A., Mathalon, D., Preda, A., Pearlson, G., ... & Calhoun, V. D. (2016). Higher dimensional meta-state analysis reveals reduced resting fMRI connectivity dynamism in schizophrenia patients. *PloS one*, *11*(3), e0149849.

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., & Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, *24*(3), 663-676.

Smith, Stephen M., et al. "Resting-state fMRI in the human connectome project." *Neuroimage* 80 (2013): 144-168.

Barch, Deanna M., et al. "Function in the human connectome: task-fMRI and individual differences in behavior." *Neuroimage* 80 (2013): 169-189.

Buysse, Daniel J., et al. "The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research." *Psychiatry research* 28.2 (1989): 193-213.

Folstein, Marshal F., Lee N. Robins, and John E. Helzer. "The mini-mental state examination." *Archives of general psychiatry*40.7 (1983): 812-812.

Gershon, Richard C., et al. "Assessment of neurological and behavioural function: the NIH Toolbox." *The Lancet Neurology*9.2 (2010): 138-139.

Achenbach, Thomas M. *The Achenbach system of empirically based assessment (ASEBA): Development, findings, theory, and applications*. University of Vermont, Research Center for Children, Youth, & Families, 2009.

Bucholz, Kathleen K., et al. "A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA." *Journal of studies on alcohol* 55.2 (1994): 149-158.

Von Luxburg, Ulrike. "A tutorial on spectral clustering." *Statistics and computing* 17.4 (2007): 395-416.

Monti, Stefano, et al. "Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data." *Machine learning* 52.1-2 (2003): 91-118.

Ṣenbabaoğlu, Yasin, George Michailidis, and Jun Z. Li. "Critical limitations of consensus clustering in class discovery." *Scientific reports* 4 (2014): 6207.

Bishop, Christopher M. "Bayesian pca." *Advances in neural information processing systems*. 1999.

Smith, Stephen M., et al. "A positive-negative mode of population covariation links brain connectivity, demographics and behavior." *Nature neuroscience* 18.11 (2015): 1565.

Gu, Q., Li, Z., & Han, J. (2012). Generalized fisher score for feature selection. *arXiv preprint arXiv:1202.3725*.

Krishnan, A., Williams, L. J., McIntosh, A. R., & Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage*, *56*(2), 455-475.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, *40*(1), 33-51.

Zöller, D., Schaer, M., Scariati, E., Padula, M. C., Eliez, S., & Van De Ville, D. (2017). Disentangling resting-state BOLD variability and PCC functional connectivity in 22q11. 2 deletion syndrome. *NeuroImage*, *149*, 85-97.

Patel, A. X., Kundu, P., Rubinov, M., Jones, P. S., Vértes, P. E., Ersche, K. D., ... & Bullmore, E. T. (2014). A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *Neuroimage*, *95*, 287-304.

Yang, Z., Zhuang, X., Screenivasan, K. R., Mishra, V. R., & Cordes, D. (2019). Robust motion regression of resting-state data using a convolutional neural network model. *Frontiers in Neuroscience*, *13*, 169.

Vidaurre, D., Smith, S. M., & Woolrich, M. W. (2017). Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*, *114*(48), 12827-12832.

Bolton, T. A., Tarun, A., Sterpenich, V., Schwartz, S., & Van De Ville, D. (2018). Interactions between large-scale functional brain networks are captured by sparse coupled HMMs. *IEEE transactions on medical imaging*, *37*(1), 230-240.

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature neuroscience*, *19*(1), 165.

Hsu, W. T., Rosenberg, M. D., Scheinost, D., Constable, R. T., & Chun, M. M. (2018). Resting-state functional connectivity predicts neuroticism and extraversion in novel individuals. *Social cognitive and affective neuroscience*, *13*(2), 224-232.

Patriat, R., Reynolds, R. C., & Birn, R. M. (2017). An improved model of motion-related signal changes in fMRI. *Neuroimage*, *144*, 74-82.

Patriat, R., Molloy, E. K., & Birn, R. M. (2015). Using edge voxel information to improve motion regression for rs-fMRI connectivity studies. *Brain connectivity*, *5*(9), 582-595.

Pruim, R. H., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage*, *112*, 267-277.

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M. (2014). Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage*, *90*, 449-468.

Bianciardi, M., Fukunaga, M., van Gelderen, P., Horovitz, S. G., de Zwart, J. A., Shmueli, K., & Duyn, J. H. (2009). Sources of functional magnetic resonance imaging signal fluctuations in the human brain at rest: a 7 T study. *Magnetic resonance imaging*, *27*(8), 1019-1029.

Liu, T. T. (2016). Noise contributions to the fMRI signal: an overview. *NeuroImage*, *143*, 141-151.

Birn, R. M. (2012). The role of physiological noise in resting-state functional connectivity. *Neuroimage*, *62*(2), 864-870.

Power, J. D., Silver, B. M., Dubin, M. J., Martin, A., & Jones, R. M. (2019). Distinctions among real and apparent respiratory motions in human fMRI data. *bioRxiv*, 601286.

Power, J. D., Plitt, M., Gotts, S. J., Kundu, P., Voon, V., Bandettini, P. A., & Martin, A. (2018). Ridding fMRI data of motion-related influences: Removal of signals with distinct spatial and physical bases in multiecho data. *Proceedings of the National Academy of Sciences*, *115*(9), E2105-E2114.

Figure legends

Figure 1: meta-state characterization on scrubbed frames. (A) Top 50 meta-states, sorted in decreasing median occurrence across subjects. Red/blue denotes significant positive/negative motion, while white highlights non-significant changes. **(B)** Across all 729 meta-states (sorted in decreasing median occurrence across subjects), percentage of subjects for which a meta-state showed significant occurrences as compared to null data. The two vertical brown lines highlight two transitions across meta-state subtypes. **(C)** For each of the 6 motion parameters, cumulative distribution of meta-states (sorted in decreasing median occurrence across subjects) showing significant positive (top plot) or negative (bottom plot) motion excursions. The dashed lines indicate the expected rise for uniform presence regardless of meta-state occurrences. **(D)** Across all (non-sorted) meta-states (top matrix), or the top 50 meta-states (sorted in decreasing median occurrence across subjects), mean transition probability from start meta-state (rows) to end meta-state (columns). Color coding denotes transition probability percentages. **(E)** Distribution, across subjects, of metric values for distance traveled in meta-state space (left boxplot) and number of visited meta-states (right boxplot).

Figure 2: groups of spatiotemporal movers on non-scrubbed frames. (A) Dimensionally reduced representation of all 224 subjects, each depicted by a three-dimensional box. Box width along the first, second and third dimension are proportional to the average motion extent, across all 6 considered time bins, in the X, Y and Z directions. Color coding in RGB scale is proportional to the extent of motion in the α (red), β (green) and γ (blue) rotational planes. Edge thickness of the boxes is proportional to the slope of a linear fit to average spatial motion over the 6 temporal bins, while red/blue symbolize increased/decreased motion over time. **(B)** Similar representation, with colors denoting the four different subgroups of movers. **(C)** Simplified representation of the data along time and clusters (top row), or along space and clusters (bottom row). Error bars denote SEM.

Figure 3: univariate links between spatiotemporal motion and anthropometry/behavior/psychometrics on non-scrubbed frames. For all 45 considered scores, Fisher score in terms of discriminability across the four spatiotemporal mover groups. Horizontal bars denote significance thresholds, derived non-parametrically and Bonferroni-corrected for 45 tests.

Figure 4: multivariate links between spatiotemporal motion and anthropometry/behavior/psychometrics on non-scrubbed frames. Z-score reflecting the importance of the 45 assessed anthropometric/behavioral/psychometric domains (left) and the 36 spatiotemporal motion features (right) for the first (A), second (B) and third (C) components from a PLS analysis (all statistically significant). Dashed horizontal lines denote the significance threshold ($|z|>3$), and text labels are appended to the domain scores showing significance. t_1 to t_6 represent the first to sixth temporal bins of a session.

Figure 5: multivariate links between global motion features and anthropometry/behavior/psychometrics on scrubbed frames. Z-score reflecting the importance of the 45 assessed anthropometric/behavioral/psychometric domains (left) and the 2 investigated global motion features (right) for the first (A) and second (B) components from a PLS analysis (both statistically significant). Dashed horizontal lines denote the significance threshold (|z|>3), and text labels are appended to the domain scores showing significance. Dist: total distance traveled in meta-state space; n_{States}: number of visited meta-states.