# Supplementary Material

## Which Gait Tasks Produce Reliable Outcome Measures of Freezing of Gait in Parkinson's Disease?

### Supplementary Material 1. Additional information on video annotation

Video annotation was performed at the two sites separately, as described in the main article. Several fidelity checks were performed to align the ratings between the centers, during which the interpretation of FOG-definitions was discussed based on several examples. An overview of the definitions used for annotation, is provided in Supplementary Table 1. Episodes of akinetic FOG, trembling FOG and festination were combined as FOG, and the impact of movement interruptions was analyzed separately using a sensitivity analysis (Supplementary Material 4).

### Supplementary Table 1. Definitions for annotation of FOG and tasks

| | | |
|---|---|---|
| **Manifestations of FOG** | | |
| Akinetic FOG | Onset | The moment the intention to move is first observed[a] and the participant is unable to do so, showing 'clear sticking of the feet' without considerable trembling movements in the legs. |
| | Termination | The moment of initial toe-off after the FOG when the participant is again able to perform at least two effective alternating steps with both legs showing no FOG-related features. |
| Trembling FOG | Onset | The moment when the foot of the participant is suddenly no longer producing an effective step forward and is displaying trembling in the legs, despite the participant's intention to continue walking, or the moment the intention to move[2] is first observed and the participant is unable to do so, showing clear trembling in the legs. |
| | Termination | The moment of initial toe-off after the FOG when the participant is again able to perform at least two effective alternating steps with both legs showing no FOG-related features. |
| Festination | Onset | The first moment of toe-off when an abnormal and high-pace oscillatory stepping behavior is observed without considerable FOG-related features. |
| | Termination | The moment of initial toe-off when the participant is again able to perform at least two effective alternating steps with both legs showing no FOG- or festination related features. |
| Movement interruptions[b] | Onset | The moment a movement interruption is first observed - when the foot of the participant is not or is suddenly no longer producing an effective step forward (despite the task instruction to do so), without a definite indication to consider it either as FOG or as voluntary stopping. |
| | Termination | The moment of initial toe-off after the movement interruption when the participant is again able to perform at least two effective alternating steps with both legs showing no FOG-related features or the first moment a definite indication is observed to annotate it as FOG or voluntary stopping. |
| **Task duration** | | |
| For calculation %TF | Onset | The moment the first intention to start the task is observed |
| | Termination | The moment the participant has finished the task |

FOG, freezing of gait; %TF, percentage time frozen. This table was adapted from[1,2] (definitions were unchanged). [a]An all-encompassing definition for the 'first intention to move' cannot be provided as this varies between situations and is therefore left for the interpretation of the expert rater. [b]Movement interruptions were not included in FOG and subtracted from the task duration; the impact of these episodes was tested separately in a sensitivity analysis.

**Supplementary Material 2. Additional information of statistical analysis**

In Supplementary Table 2, an overview is provided of the outcomes tested using (Generalized) Linear Mixed Models ((G)LMMs). In the GLMMs, a binary distribution and logit link function was used. In all models, an unstructured covariance matrix was used if possible. Only for the consistency of the presence of FOG, the model would not converge and a first-order autoregressive structure was selected instead after checking the model fit (Generalized Chi Square / degrees of freedom). The models were applied first to compare differences in OFF vs. ON and then to compare OFF+ON with OFF.

Cohen's d between-group (equation 1) and within-group (equation 2) effect sizes were calculated based on t-statistics and degrees of freedom (Satterthwaite correction) from the model output.[3] For main and interaction effects, F-statistics (type III) were first converted to t-statistics (equation 3). As this provides absolute values, we reported all effect sizes as absolute values for consistency.

$$d_{between-group} = \frac{2 \cdot t_{statistic}}{\sqrt{(df+2)}} \ (equation\ 1), \qquad d_{within-group} = \frac{t_{statistic}}{\sqrt{(df+1)}} \ (equation\ 2),$$

$$t_{statistic} = \sqrt{F_{statistic}} \ (equation\ 3)$$

**Supplementary Table 2. An overview of the different outcomes in the statistical models**

| | Model | Outcome sensitivity (T1, N=63) | | Outcome reliability (T1 vs. T2, N=26) | |
|---|---|---|---|---|---|
| | | Outcome | Shown on group level as | Outcome | Shown on group level as |
| FOG-presence | GLMM | Presence of FOG (binary: 0=no, 1=yes) | %N with FOG | Consistency presence of FOG (binary: 0=no, 1=yes) | %N with consistent FOG |
| FOG-severity | LMM | %TF (continuous) | mean (SD) %TF | Δ%TF (T2-T1) (continuous) | mean (SD) Δ%TF |

T1, timepoint 1; T2, timepoint 2; FOG, freezing of gait; %TF, percentage time frozen; GLMM, Generalized Linear Mixed Model; LMM, Linear Mixed Model; %N, percentage of people relative to the total; SD, standard deviation; Δ%TF, $\%TF_{T2}$-$\%TF_{T1}$.

**Supplementary Material 3. Overview results mixed models**

*Main and interaction effects statistical models*

In the main article, only the most important results of the statistical models were reported. Here, we provide a more complete overview. Main and interaction effects (type III F-tests) are shown in Supplementary Table 3. In the sections below, post-hoc results are shown in the tables together with the raw data for interpretation.

**Supplementary Table 3. Main and interaction effects statistical models.**

| | FOG-presence (binary) | | FOG-severity (%TF) | |
|---|---|---|---|---|
| | **Sensitivity** | **Test-retest reliability** | **Sensitivity** | **Test-retest reliability** |
| **Medication=OFF, ON** | | | | |
| Medication | d=0.70; **p<0.001*** | d=0.20; p=0.317 | d=0.74; **p<0.001*** | d=0.06; p=0.782 |
| Task | d=0.49; **p<0.001*** | d=0.12; **p=0.043*** | d=0.67; **p<0.001*** | d=0.20; p=0.447 |
| Medication*task | d=0.10; p=0.684 | d=0.10; p=0.234 | d=0.20; **p=0.026*** | d=0.22; p=0.333 |
| Center | d=0.04; p=0.862 | d=0.39; p=0.335 | d=0.02; p=0.944 | d=0.27; p=0.505 |
| **Medication=OFF+ON, OFF** | | | | |
| Medication | d=0.46; **p<0.001*** | d=0.10; p=0.626 | d=0.19; p=0.150 | d=0.10; p=0.616 |
| Task | d=0.43; **p<0.001*** | d=0.16; **p=0.003*** | d=0.66; **p<0.001*** | d=0.14; p=0.808 |
| Medication*task | d=0.17; p=0.117 | d=0.03; p=0.997 | d=0.20; **p=0.038*** | d=0.29; p=0.114 |
| Center | d=0.01; p=0.980 | d=0.41; p=0.322 | d=0.08; p=0.745 | d=0.28; p=0.493 |

GLMM, Generalized Linear Mixed Model; LMM, Linear Mixed Model; %TF, percentage time frozen; Δ%TF, $\%TF_{T2}$-$\%TF_{T1}$; OFF, >12 h withdrawal of dopaminergic medication; ON, 1 h after intake of dopaminergic medication; OFF+ON, average OFF and ON; d, Cohen's d (absolute); p, p-value.

*Post-hoc comparisons for sensitivity of FOG-presence*

Supplementary Table 4 shows task comparisons for sensitivity of FOG-presence, together with the raw data for interpretation (T1, N=63). Results are shown first for the model including OFF and ON, and also for the model including OFF+ON and OFF.

**Supplementary Table 4. Post-hoc task comparisons for sensitivity of FOG-presence.**

| | 4MW | TUG ST | TUG DT | 360° turns ST | 360° turns DT | Hotspot door | Personalized hotspot |
|---|---|---|---|---|---|---|---|
| **Raw data** | | | | | | | |
| %N with FOG, OFF+ON | 44.4 | 60.3 | 66.7 | 95.2 | 95.2 | 82.5 | 95.2 |
| %N with FOG, OFF | 41.3 | 54.0 | 63.5 | 88.9 | 92.1 | 69.8 | 88.7 |
| %N with FOG, ON | 20.6 | 33.3 | 39.7 | 77.4 | 75.8 | 46.0 | 58.7 |
| **Task comparisons, with medication= OFF, ON** | | | | | | | |
| 4MW | | d=0.36; p=0.082 | d=0.49; **p=0.005*** | d=0.94; **p<0.001*** | d=0.91; **p<0.001*** | d=0.59; **p<0.001*** | d=0.85; **p<0.001*** |
| TUG ST | d=0.36; p=0.082 | | d=0.23; p=0.569 | d=0.79; **p<0.001*** | d=0.78; **p<0.001*** | d=0.38; p=0.055 | d=0.67; **p<0.001*** |
| TUG DT | d=0.49; **p=0.005*** | d=0.23; p=0.569 | | d=0.66; **p<0.001*** | d=0.67; **p<0.001*** | d=0.17; p=0.850 | d=0.48; **p=0.007*** |
| 360° turns ST | d=0.94; **p<0.001*** | d=0.79; **p<0.001*** | d=0.66; **p<0.001*** | | d=0.04; p=1.000 | d=0.50; **p=0.004*** | d=0.17; p=0.850 |
| 360° turns DT | d=0.91; **p<0.001** | d=0.78; **p<0.001*** | d=0.67; **p<0.001*** | d=0.04; p=1.00 | | d=0.58; **p<0.001*** | d=0.20; p=0.709 |
| Hotspot door | d=0.59; **p<0.001*** | d=0.38; p=0.055 | d=0.17; p=0.850 | d=0.50; **p=0.004*** | d=0.58; **p<0.001*** | | d=0.40; **p=0.042*** |
| Personalized hotspot | d=0.85; **p<0.001*** | d=0.67; **p<0.001*** | d=0.48; **p=0.007*** | d=0.17; p=0.850 | d=0.20; p=0.709 | d=0.40; **p=0.042*** | |
| **Task comparisons, with medication= OFF+ON, OFF** | | | | | | | |
| 4MW | | d=0.31; p=0.207 | d=0.43; **p=0.021*** | d=0.72; **p<0.001*** | d=0.71; **p<0.001*** | d=0.59; **p<0.001*** | d=0.71; **p<0.001*** |
| TUG ST | d=0.31; p=0.207 | | d=0.19; p=0.740 | d=0.55; **p=0.001*** | d=0.58; **p<0.001*** | d=0.42; **p=0.027*** | d=0.59; **p<0.001*** |
| TUG DT | d=0.43; **p=0.021*** | d=0.19; p=0.740 | | d=0.48; **p=0.007*** | d=0.51; **p=0.003*** | d=0.25; p=0.422 | d=0.47; **p=0.007*** |
| 360° turns ST | d=0.72; **p<0.001*** | d=0.55; **p=0.001*** | d=0.48; **p=0.007*** | | d=0.03; p=1.000 | d=0.31; p=0.216 | d=0.00; p=1.000 |
| 360° turns DT | d=0.71; **p<0.001*** | d=0.58; **p<0.001*** | d=0.51; **p=0.003*** | d=0.03; p=1.000 | | d=0.38; p=0.059 | d=0.04; p=1.000 |
| Hotspot door | d=0.59; **p<0.001*** | d=0.42; **p=0.027*** | d=0.25; p=0.422 | d=0.31; p=0.216 | d=0.38; p=0.059 | | d=0.32; p=0.162 |
| Personalized hotspot | d=0.71; **p<0.001*** | d=0.59; **p<0.001*** | d=0.47; **p=0.007*** | d=0.00; p=1.000 | d=0.04; p=1.000 | d=0.32; p=0.162 | |

4MW, Four meter walk; TUG, Timed Up and Go; ST, single task; DT, dual task; %N, percentage of people relative to the total; OFF, >12 h withdrawal of dopaminergic medication; ON, 1 h after intake of dopaminergic medication; OFF+ON, average OFF and ON; d, Cohen's d (absolute); p, p-value. * significant difference after Tukey-Kramer adjustment.

*Post-hoc comparisons for test-retest reliability of FOG-presence*

Supplementary Table 5 shows task comparisons for consistency presence of FOG, together with the raw data for interpretation. Results are shown first for the model including OFF and ON, and also for the model including OFF+ON and OFF.

**Supplementary Table 5. Post-hoc task comparisons for test-retest reliability of FOG-presence.**

| | 4MW | TUG ST | TUG DT | 360° turns ST | 360° turns DT | Hotspot door | Personalized hotspot |
|---|---|---|---|---|---|---|---|
| **Raw data** | | | | | | | |
| %N with consistent FOG OFF+ON | 72.0 | 68.0 | 92.0 | 87.5 | 95.7 | 66.7 | 75.0 |
| %N with consistent FOG OFF | 68.0 | 60.0 | 92.0 | 80.0 | 95.8 | 66.7 | 66.7 |
| %N with consistent FOG ON | 80.8 | 61.5 | 73.1 | 76.0 | 76.0 | 65.4 | 80.8 |
| **Task comparisons, with medication= OFF, ON** | | | | | | | |
| 4MW | | d=0.13; p=0.700 | d=0.09; p=0.915 | d=0.03; p=1.000 | d=0.13; p=0.695 | d=0.08; p=0.961 | d=0.01; p=1.000 |
| TUG ST | d=0.13; p=0.700 | | d=0.21; p=0.142 | d=0.15; p=0.514 | d=0.22; p=0.117 | d=0.04; p=0.998 | d=0.12; p=0.792 |
| TUG DT | d=0.09; p=0.915 | d=0.21; p=0.142 | | d=0.07; p=0.980 | d=0.05; p=0.997 | d=0.16; p=0.427 | d=0.10; p=0.899 |
| 360° turns ST | d=0.03; p=1.000 | d=0.15; p=0.514 | d=0.07; p=0.980 | | d=0.11; p=0.823 | d=0.11; p=0.835 | d=0.04; p=1.000 |
| 360° turns DT | d=0.13; p=0.695 | d=0.22; p=0.117 | d=0.05; p=0.997 | d=0.11; p=0.823 | | d=0.19; p=0.210 | d=0.13; p=0.663 |
| Hotspot door | d=0.08; p=0.961 | d=0.04; p=0.998 | d=0.16; p=0.427 | d=0.11; p=0.835 | d=0.19; p=0.210 | | d=0.08; p=0.964 |
| Personalized hotspot | d=0.01; p=1.000 | d=0.12; p=0.792 | d=0.10; p=0.899 | d=0.04; p=1.000 | d=0.13; p=0.663 | d=0.08; p=0.964 | |
| **Task comparisons, with medication= OFF+ON, OFF** | | | | | | | |
| 4MW | | d=0.06; p=0.994 | d=0.22; p=0.136 | d=0.13; p=0.736 | d=0.24; p=0.083 | d=0.03; p=1.000 | d=0.01; p=1.000 |
| TUG ST | d=0.06; p=0.994 | | d=0.28; **p=0.020*** | d=0.18; p=0.357 | d=0.27; **p=0.032*** | d=0.02; p=1.000 | d=0.06; p=0.990 |
| TUG DT | d=0.22; p=0.136 | d=0.28; **p=0.020*** | | d=0.11; p=0.825 | d=0.06; p=0.989 | d=0.24; p=0.083 | d=0.21; p=0.190 |
| 360° turns ST | d=0.13; p=0.736 | d=0.18; p=0.357 | d=0.11; p=0.825 | | d=0.16; p=0.467 | d=0.15; p=0.531 | d=0.12; p=0.811 |
| 360° turns DT | d=0.24; p=0.083 | d=0.27; **p=0.032*** | d=0.06; p=0.989 | d=0.16; p=0.467 | | d=0.27; **p=0.027*** | d=0.23; p=0.094 |
| Hotspot door | d=0.03; p=1.000 | d=0.02; p=1.000 | d=0.24; p=0.083 | d=0.15; p=0.531 | d=0.27; **p=0.027*** | | d=0.04; p=0.999 |
| Personalized hotspot | d=0.01; p=1.000 | d=0.06; p=0.990 | d=0.21; p=0.190 | d=0.12; p=0.811 | d=0.23; p=0.094 | d=0.04; p=0.999 | |

4MW, Four meter walk; TUG, Timed Up and Go; ST, single task; DT, dual task; %N, percentage of people relative to the total; OFF, >12 h withdrawal of dopaminergic medication; ON, 1 h after intake of dopaminergic medication; OFF+ON, average OFF and ON; d, Cohen's d (absolute); p, p-value. * significant difference after Tukey-Kramer adjustment.

*Post-hoc comparisons for sensitivity of FOG-severity*

Supplementary Table 6 (medication = OFF vs. ON) and Supplementary Table 7 (medication=OFF+ON vs. OFF) present the post-hoc results of the interaction between the task and medication state for %TF. The tables show the raw data, the medication effects per task, and all task comparisons per medication state.

**Supplementary Table 6. Post-hoc results task*medication for sensitivity of FOG-severity (medication = OFF vs. ON)**

| | | 4MW | TUG ST | TUG DT | 360° turns ST | 360° turns DT | Hotspot door | Personalized hotspot |
|---|---|---|---|---|---|---|---|---|
| **Raw data** | | | | | | | | |
| %TF OFF, mean (SD) | | 6.83 (13.5) | 9.67 (15.2) | 14.8 (21.9) | 40.0 (34.1) | 46.2 (34.6) | 16.1 (17.8) | 19.2 (17.8) |
| %TF ON, mean (SD) | | 3.09 (9.35) | 3.28 (6.89) | 4.71 (10.3) | 29.7 (31.1) | 31.8 (34.1) | 9.50 (15.9) | 6.67 (11.5) |
| **Medication effects per task** | | | | | | | | |
| OFF vs. ON | | d=0.46; **p=0.032*** | d=0.54; **p=0.005*** | d=0.62; **p=0.001*** | d=0.28; p=0.632 | d=0.47; **p=0.028*** | d=0.40; p=0.111 | d=0.71; **p<0.001*** |
| **Task comparisons in OFF and ON** | | | | | | | | |
| OFF | 4MW | | d=0.29; p=0.597 | d=0.48; **p=0.023*** | d=1.22; **p<0.001*** | d=1.41; **p<0.001*** | d=0.65; **p<0.001*** | d=1.03; **p<0.001*** |
| | TUG ST | d=0.29; p=0.597 | | d=0.36; p=0.242 | d=0.93; **p<0.001*** | d=1.15; **p<0.001*** | d=0.46; **p=0.036*** | d=0.80; **p<0.001*** |
| | TUG DT | d=0.48; **p=0.023*** | d=0.36; p=0.242 | | d=0.74; **p<0.001*** | d=0.94; **p<0.001*** | d=0.23; p=0.857 | d=0.51; **p=0.010*** |
| | 360° turns ST | d=1.22; **p<0.001*** | d=0.93; **p<0.001*** | d=0.74; **p<0.001*** | | d=0.22; p=0.898 | d=0.55; **p=0.004*** | d=0.35; p=0.286 |
| | 360° turns DT | d=1.41; **p<0.001*** | d=1.15; **p<0.001*** | d=0.94; **p<0.001*** | d=0.22; p=0.898 | | d=0.74; **p<0.001*** | d=0.54; **p=0.006*** |
| | Hotspot door | d=0.65; **p<0.001*** | d=0.46; **p=0.036*** | d=0.23; p=0.857 | d=0.55; **p=0.004*** | d=0.74; **p<0.001*** | | d=0.33; p=0.374 |
| | Personalized hotspot | d=1.03; **p<0.001*** | d=0.80; **p<0.001*** | d=0.51; **p=0.010*** | d=0.35; p=0.286 | d=0.54; **p=0.006*** | d=0.33; p=0.374 | |
| ON | 4MW | | d=0.23; p=0.868 | d=0.34; p=0.329 | d=1.22; **p<0.001*** | d=1.14; **p<0.001*** | d=0.51; **p=0.010*** | d=0.60; **p=0.001*** |
| | TUG ST | d=0.23; p=0.868 | | d=0.17 p=0.989 | d=1.27; **p<0.001*** | d=1.15; **p<0.001*** | d=0.41; p=0.095 | d=0.48; **p=0.022*** |
| | TUG DT | d=0.34; p=0.329 | d=0.17; p=0.989 | | d=1.12; **p<0.001*** | d=1.07; **p<0.001*** | d=0.29; p=0.592 | d=0.34; p=0.308 |
| | 360° turns ST | d=1.22; **p<0.001*** | d=1.27; **p<0.001*** | d=1.12; **p<0.001*** | | d=0.03; p=1.000 | d=0.75; **p<0.001*** | d=0.74; **p<0.001*** |
| | 360° turns DT | d=1.14; **p<0.001*** | d=1.15; **p<0.001*** | d=1.07; **p<0.001*** | d=0.03; p=1.000 | | d=0.71; **p<0.001*** | d=0.70; **p<0.001*** |
| | Hotspot door | d=0.51; **p=0.010*** | d=0.41; p=0.095 | d=0.29; p=0.592 | d=0.75; **p<0.001*** | d=0.71; **p<0.001*** | | d=0.04; p=1.000 |
| | Personalized hotspot | d=0.60; **p=0.001*** | d=0.48; **p=0.022*** | d=0.34; p=0.308 | d=0.74; **p<0.001*** | d=0.70; **p<0.001*** | d=0.04; p=1.000 | |

4MW, Four meter walk; TUG, Timed Up and Go; ST, single task; DT, dual task; OFF, >12 h withdrawal of dopaminergic medication; ON, 1 h after intake of dopaminergic medication; %TF, percentage time frozen; SD, standard deviation; d, Cohen's d (absolute); p, p-value. * significant difference after Tukey-Kramer adjustment. Statistical results are reported based on the inverse hyperbolic sine transformation of %TF.

## Supplementary Table 7. Post-hoc results task*medication for sensitivity of FOG-severity (medication = OFF+ON vs. OFF)

| | | 4MW | TUG ST | TUG DT | 360° turns ST | 360° turns DT | Hotspot door | Personalized hotspot |
|---|---|---|---|---|---|---|---|---|
| **Raw data** | | | | | | | | |
| %TF OFF+ON, mean (SD) | | 4.96 (9.64) | 6.48 (9.76) | 9.78 (13.9) | 34.4 (28.6) | 38.6 (30.1) | 12.8 (13.6) | 12.9 (12.6) |
| %TF OFF, mean (SD) | | 6.83 (13.5) | 9.67 (15.2) | 14.8 (21.9) | 40.0 (34.1) | 46.2 (34.6) | 16.1 (17.8) | 19.2 (17.8) |
| **Medication effects per task** | | | | | | | | |
| OFF+ON vs. OFF | | d=0.19; p=0.961 | d=0.28; p=0.615 | d=0.37; p=0.210 | d=0.10; p=1.000 | d=0.11; p=1.000 | d=0.04; p=1.000 | d=0.26; p=0.728 |
| **Task comparisons in OFF+ON and OFF** | | | | | | | | |
| OFF+ON | 4MW | | d=0.34; p=0.334 | d=0.51; **p=0.012*** | d=1.58; **p<0.001*** | d=1.62; **p<0.001*** | d=0.87; **p<0.001*** | d=1.21; **p<0.001*** |
| | TUG ST | d=0.34; p=0.334 | | d=0.37; p=0.206 | d=1.28; **p<0.001*** | d=1.43; **p<0.001*** | d=0.65; **p<0.001*** | d=0.97; **p<0.001*** |
| | TUG DT | d=0.51; **p=0.012*** | d=0.37; p=0.206 | | d=1.07; **p<0.001*** | d=1.20; **p<0.001*** | d=0.41; p=0.097 | d=0.61; **p<0.001*** |
| | 360° turns ST | d=1.58; **p<0.001*** | d=1.28; **p<0.001*** | d=1.07; **p<0.001*** | | d=0.14; p=0.998 | d=0.72; **p<0.001*** | d=0.65; **p<0.001*** |
| | 360° turns DT | d=1.62; **p<0.001*** | d=1.43; **p<0.001*** | d=1.20; **p<0.001*** | d=0.14; p=0.998 | | d=0.84; **p<0.001*** | d=0.77; **p<0.001*** |
| | Hotspot door | d=0.87; **p<0.001*** | d=0.65; **p<0.001*** | d=0.41; p=0.097 | d=0.72; **p<0.001*** | d=0.84; **p<0.001*** | | d=0.25; p=0.803 |
| | Personalized hotspot | d=1.21; **p<0.001*** | d=0.97; **p<0.001*** | d=0.61; **p<0.001*** | d=0.65; **p<0.001*** | d=0.77; **p<0.001*** | d=0.25; p=0.803 | |
| OFF | 4MW | | d=0.29; p=0.597 | d=0.48; **p=0.023*** | d=1.22; **p<0.001*** | d=1.41; **p<0.001*** | d=0.65; **p<0.001*** | d=1.02; **p<0.001*** |
| | TUG ST | d=0.29; p=0.597 | | d=0.36; p=0.242 | d=0.93; **p<0.001*** | d=1.15; **p<0.001*** | d=0.46; **p=0.036*** | d=0.80; **p<0.001*** |
| | TUG DT | d=0.48; **p=0.023*** | d=0.36; p=0.242 | | d=0.74; **p<0.001*** | d=0.94; **p<0.001*** | d=0.23; p=0.857 | d=0.52; **p=0.010*** |
| | 360° turns ST | d=1.22; **p<0.001*** | d=0.93; **p<0.001*** | d=0.74; **p<0.001*** | | d=0.22; p=0.898 | d=0.55; **p=0.004*** | d=0.34; p=0.290 |
| | 360° turns DT | d=1.41; **p<0.001*** | d=1.15; **p<0.001*** | d=0.94; **p<0.001*** | d=0.22; p=0.898 | | d=0.74; **p<0.001*** | d=0.53; **p=0.006*** |
| | Hotspot door | d=0.65; **p<0.001*** | d=0.46; **p=0.036*** | d=0.23; p=0.857 | d=0.55; **p=0.004*** | d=0.74; **p<0.001*** | | d=0.33; p=0.367 |
| | Personalized hotspot | d=1.02; **p<0.001*** | d=0.80; **p<0.001*** | d=0.52; **p=0.010*** | d=0.34; p=0.290 | d=0.53; **p=0.006*** | d=0.33; p=0.367 | |

4MW, Four meter walk; TUG, Timed Up and Go; ST, single task; DT, dual task; OFF, >12 h withdrawal of dopaminergic medication; ON, 1 h after intake of dopaminergic medication; %TF, percentage time frozen; SD, standard deviation. d, Cohen's d (absolute); p, p-value. * significant difference after Tukey-Kramer adjustment. Statistical results are reported based on the inverse hyperbolic sine transformation of %TF.

**Supplementary Material 4. Testing the impact of movement interruptions**

*Methods*

Movement interruptions are episodes which we could not identify with certainty as FOG or as a voluntary stop,[1] mostly because we could not observe on the video whether there was an intention to move. This could occur during dual tasking; if the patient stopped while counting, it could be unclear whether the patient was unable to move forward due to the additional cognitive load (FOG), or whether the patient prioritized the cognitive task over the motor task (voluntary stop). However, this could also occur due to other reasons, for instance if we were unable to observe an attempt to move due to pure akinesia or (environmental) distractions.

We analyzed movement interruptions separately, and performed a sensitivity analysis to determine whether including them influenced the outcomes. The (G)LMM analyses were repeated with and without including movement interruptions in FOG.

*Results*

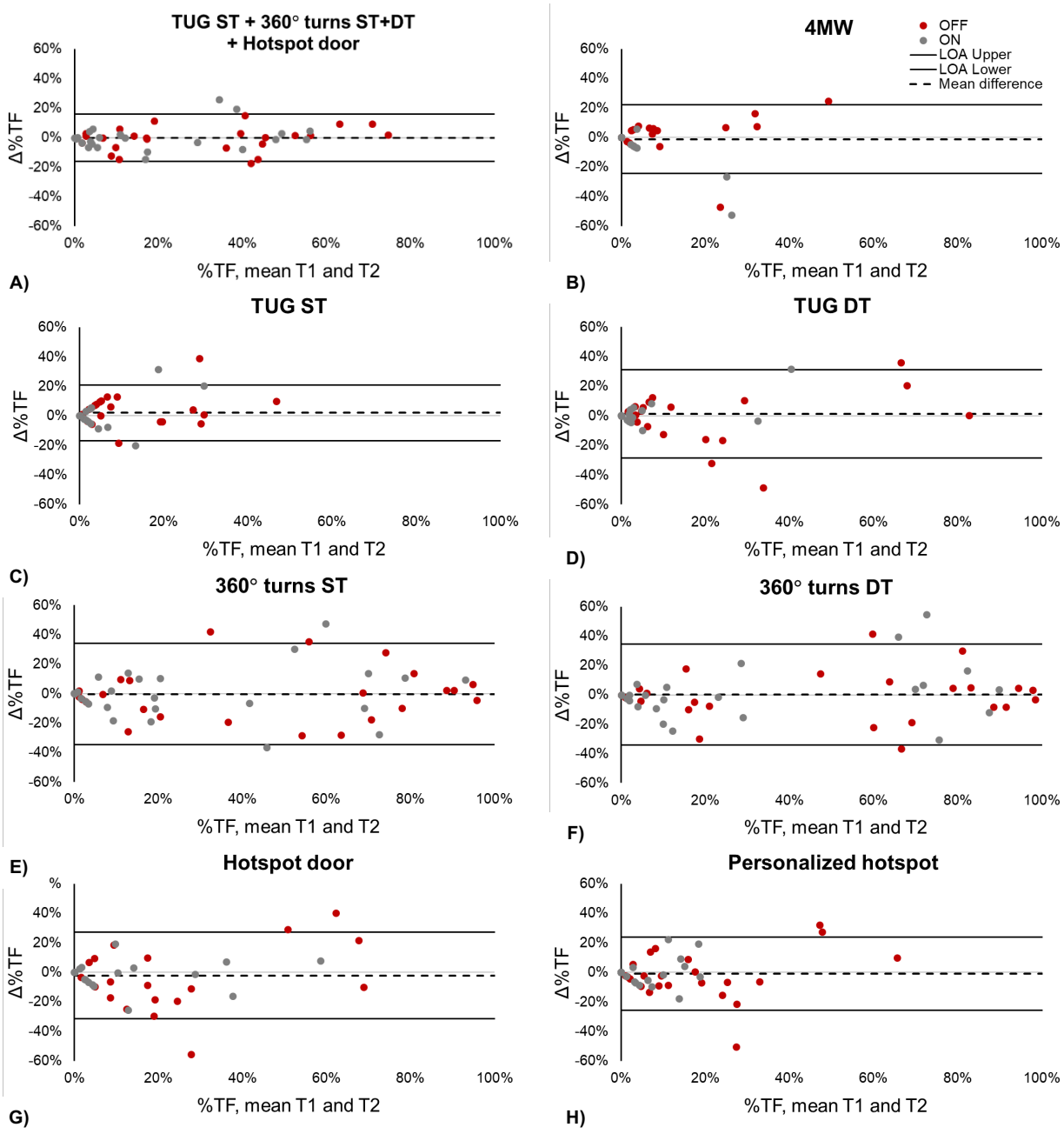
%TF of movement interruptions was very low (<4% in all conditions). Still, about a third of patients had experienced some movement interruption during 360° turns DT at baseline (Supplementary Table 8). However, outcomes were very similar between FOG with and without movement interruptions. For FOG with movement interruptions, %N with FOG was slightly higher for Hotspot door, influencing the difference with TUG ST (now significant: d=0.41, p=0.028), and the Personalized hotspot (no longer significant: p=0.105). In addition, the medication effect on %TF for Hotspot door became significant (d=0.45; p=0.044). We further found in OFF that %TF TUG DT was significantly higher than TUG ST when including movement interruptions (d=0.47, p=0.031), but not without them (p=0.242). For 360° turns, ST-DT differences were not affected by inclusion of movement interruptions (still not significant).

**Supplementary Table 8. Sensitivity of movement interruptions, and sensitivity and test-retest reliability for FOG without and with movement interruptions**

| | | 4MW | TUG ST | TUG DT | 360° turns ST | 360° turns DT | Hotspot door | Personal. hotspot |
|---|---|---|---|---|---|---|---|---|
| **Sensitivity of movement interruptions[1]** | | | | | | | | |
| %N with movement interruptions | OFF+ON | 3.17 | 3.17 | 15.9 | 19.4 | 40.3 | 11.1 | 19.4 |
| | OFF | 3.17 | 3.17 | 14.3 | 17.5 | 31.7 | 7.94 | 17.7 |
| | ON | 0.00 | 0.00 | 4.76 | 4.84 | 22.6 | 3.17 | 6.35 |
| %TF movement interruptions mean (SD) | OFF+ON | 0.09 (0.50) | 0.04 (0.22) | 1.14 (3.89) | 0.38 (0.96) | 3.11 (6.92) | 0.54 (2.07) | 0.55 (1.77) |
| | OFF | 0.17 (1.00) | 0.08 (0.44) | 1.95 (7.07) | 0.61 (1.75) | 3.87 (10.3) | 0.64 (2.67) | 0.82 (2.38) |
| | ON | 0.00 (0.00) | 0.00 (0.00) | 0.34 (1.78) | 0.14 (0.74) | 2.28 (5.42) | 0.44 (3.25) | 0.27 (1.55) |
| **Comparing FOG with and without movement interruptions** | | | | | | | | |
| ***%N with FOG[a]*** | | | | | | | | |
| FOG | OFF+ON | 44.4 | 60.3 | 66.7 | 95.2 | 95.2 | 82.5 | 95.2 |
| | OFF | 41.3 | 54.0 | 63.5 | 88.9 | 92.1 | 69.8 | 88.7 |
| | ON | 20.6 | 33.3 | 39.7 | 77.4 | 75.8 | 46.0 | 58.7 |
| FOG + movement interruptions | OFF+ON | 46.0 | 60.3 | 69.8 | 95.2 | 95.2 | 82.5 | 95.2 |
| | OFF | 42.9 | 54.0 | 66.7 | 88.9 | 92.1 | 73.0 | 88.7 |
| | ON | 20.6 | 33.3 | 41.3 | 79.0 | 77.4 | 47.6 | 61.9 |
| ***%N with consistent FOG[b]*** | | | | | | | | |
| FOG | OFF+ON | 72.0 | 68.0 | 92.0 | 87.5 | 95.7 | 66.7 | 75.0 |
| | OFF | 68.0 | 60.0 | 92.0 | 80.0 | 95.8 | 66.7 | 66.7 |
| | ON | 80.8 | 61.5 | 73.1 | 76.0 | 76.0 | 65.4 | 80.8 |
| FOG + movement interruptions | OFF+ON | 68.0 | 68.0 | 92.0 | 87.5 | 95.7 | 66.7 | 75.0 |
| | OFF | 64.0 | 60.0 | 92.0 | 84.0 | 95.8 | 70.8 | 66.7 |
| | ON | 80.8 | 57.7 | 76.9 | 80.0 | 84.0 | 65.4 | 76.9 |
| ***%TF, mean (SD)[a]*** | | | | | | | | |
| FOG | OFF+ON | 4.96 (9.64) | 6.48 (9.76) | 9.78 (13.9) | 34.4 (28.6) | 38.6 (30.1) | 12.8 (13.6) | 12.9 (12.6) |
| | OFF | 6.83 (13.5) | 9.67 (15.2) | 14.8 (21.9) | 40.0 (34.1) | 46.2 (34.6) | 16.1 (17.8) | 19.2 (17.8) |
| | ON | 3.09 (9.35) | 3.28 (6.89) | 4.71 (10.3) | 29.7 (31.1) | 31.8 (34.1) | 9.50 (15.9) | 6.67 (11.5) |
| FOG + movement interruptions | OFF+ON | 5.00 (9.68) | 6.51 (9.75) | 10.7 (14.6) | 34.6 (28.5) | 40.5 (30.4) | 13.3 (13.6) | 13.3 (12.7) |
| | OFF | 6.92 (13.6) | 9.75 (15.2) | 16.4 (22.8) | 40.3 (34.0) | 48.4 (34.4) | 16.7 (17.7) | 19.8 (18.0) |
| | ON | 3.09 (9.35) | 3.28 (6.89) | 4.94 (10.9) | 29.8 (31.1) | 33.3 (33.9) | 9.85 (16.3) | 6.94 (11.5) |
| ***ICC %TF, mean (95%CI)[b]*** | | | | | | | | |
| FOG | OFF+ON | 0.77 (0.55; 0.89) | 0.69 (0.41; 0.85) | 0.73 (0.48; 0.87) | 0.91 (0.80; 0.96) | 0.92 (0.81; 0.96) | 0.80 (0.59; 0.91) | 0.52 (0.14; 0.76) |
| | OFF | 0.66 (0.36; 0.83) | 0.71 (0.44; 0.86) | 0.79 (0.58; 0.90) | 0.88 (0.75; 0.95) | 0.90 (0.78; 0.95) | 0.66 (0.35; 0.84) | 0.63 (0.32; 0.82) |
| | ON | 0.18 (-0.19; 0.51) | 0.40 (0.02; 0.68) | 0.55 (0.23; 0.77) | 0.85 (0.68; 0.93) | 0.87 (0.73; 0.94) | 0.87 (0.73; 0.94) | 0.44 (0.07; 0.71) |
| FOG + movement interruptions | OFF+ON | 0.77 (0.54; 0.89) | 0.70 (0.43; 0.85) | 0.72 (0.47; 0.87) | 0.91 (0.80; 0.96) | 0.90 (0.78; 0.96) | 0.78 (0.56; 0.90) | 0.54 (0.17; 0.77) |
| | OFF | 0.65 (0.35; 0.83) | 0.73 (0.47; 0.87) | 0.78 (0.55; 0.89) | 0.90 (0.76; 0.95) | 0.87 (0.73; 0.94) | 0.63 (0.32; 0.82) | 0.66 (0.36; 0.84) |
| | ON | 0.18 (-0.19; 0.51) | 0.39 (0.00; 0.68) | 0.57 (0.25; 0.78) | 0.86 (0.70; 0.93) | 0.86 (0.71; 0.94) | 0.87 (0.73; 0.94) | 0.43 (0.05; 0.70) |

4MW, Four meter walk; TUG, Timed-Up and Go; ST, single task; DT, dual task; FOG, freezing of gait; SD, standard deviation; %TF, percentage time frozen; OFF, >12 h withdrawal of dopaminergic medication; ON, 1 h after intake of dopaminergic medication; OFF+ON, %TF averaged over the OFF and ON state; ICC, intra-class correlation coefficient; CI, confidence interval. Outcomes are categorized as [a] sensitivity to provoke FOG (T1, N=63), [b] test-retest reliability (T1 vs. T2, N=26).

**Supplementary Material 5. Bland-Altman plots %TF per task**




**Supplementary Figure 1. Bland-Altman plot for A) the optimal protocol and B-H) per task, in OFF (red) and ON (grey).** 4MW, 4 meter walk; TUG, Timed-Up and Go; ST, single task; DT, dual task; %TF, percentage time frozen; Δ%TF, $\%TF_{T2}$-$\%TF_{T1}$; T1, timepoint 1; T2, timepoint 2; OFF, >12 h withdrawal of dopaminergic medication; ON, 1 h after intake of dopaminergic medication; LOA, limits of agreement.

**Supplementary Material 6. Determining the optimal task combination**

*Methods*

In order to facilitate decision making in future studies, we investigated what task combination resulted in an outcome of FOG-severity (%TF) with the lowest MDC index. For each medication state, we calculated the minimal detectable change relative to the mean sensitivity (MDC index) for all task combinations for up to 4 tasks. We chose 4 tasks as the limit to obtain a protocol that is feasible in many studies. With this many combinations, there is a large likelihood that multiple combinations perform similarly well. Therefore, we did not simply select the combination with the minimal MDC index. Instead, we selected all task combinations within the lowest 5% of MDC indices, and subsequently determined the proportion of times each task was included in these combinations.

*Results*

As shown in Supplementary Table 9, 360° turns ST was included most often in the task combinations. For 2 tasks, the second task depended on the medication state (OFF+ON: TUG ST, OFF: 360° turns DT, ON: Hotspot door). For 3 tasks, the second task was TUG ST, and the third task was 360° turns DT in OFF+ON and OFF but Hotspot door in ON. For 4 tasks, the best performing combination was TUG ST+ 360° turns ST+DT+ Hotspot door in both OFF and ON. In OFF+ON, performance was slightly better when including Personalized hotspot instead of Hotspot door (MDC index=0.45 instead of 0.48). Overall, the 4 tasks that performed best in combination with others were: 360° turns ST, followed by 360° turns DT, TUG ST and Hotspot door. The combination of these 4 tasks was selected as the "optimal protocol" despite a slightly lower MDC index for 3 tasks in ON, to obtain one protocol which is reliable in both medication states. Sensitivity and test-retest reliability outcomes are reported in the main article (Table 2), and a Bland-Altman plot of the optimal protocol is presented in Supplementary Figure 1A.

**Supplementary Table 9. The proportion of times each task was included in the 5% best performing task combinations (lowest MDC index), for a combination of 2, 3, or 4 tasks.**

| | | 4MW | TUG ST | TUG DT | 360° turns ST | 360° turns DT | Hotspot door | Personalized hotspot | Lowest MDC index |
|---|---|---|---|---|---|---|---|---|---|
| 2 tasks | OFF+ON | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.55 |
| | OFF | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.57 |
| | ON | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.95 |
| | *sum* | *0.00* | *1.00* | *0.00* | *3.00* | *1.00* | *1.00* | *0.00* | |
| 3 tasks | OFF+ON | 0.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.50 |
| | OFF | 0.00 | 0.60 | 0.00 | 1.00 | 1.00 | 0.40 | 0.00 | 0.57 |
| | ON | 0.00 | 0.60 | 0.00 | 1.00 | 0.00 | 1.00 | 0.40 | 0.87 |
| | *sum* | *0.00* | *2.20* | *0.00* | *3.00* | *2.00* | *1.40* | *0.40* | |
| 4 tasks | OFF+ON | 0.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.40 | 0.60 | 0.45 |
| | OFF | 0.00 | 1.00 | 0.00 | 1.00 | 1.00 | 0.75 | 0.25 | 0.53 |
| | ON | 0.00 | 0.71 | 0.00 | 1.00 | 1.00 | 1.00 | 0.29 | 0.90 |
| | *sum* | *0.00* | *2.71* | *0.00* | *3.00* | *3.00* | *2.15* | *1.14* | |
| *Total sum* | | *0.00* | *5.91* | *0.00* | *9.00* | *6.00* | *4.55* | *1.54* | |

MDC index, Minimal detectable change/mean %TF at T1; 4MW, Four Meter Walk; TUG, Timed-Up and Go; ST, single task; DT, dual task; OFF, >12 h withdrawal of dopaminergic medication; ON, 1 h after dopaminergic medication intake. OFF+ON: averaged over OFF and ON states.

**Supplementary Material 7. Test-retest reliability of medication effects on %TF**

*Methods*

FOG-provoking protocols are not only used to reliably assess FOG-severity, but also to evaluate the effects of a particular intervention on FOG-severity. Since dopaminergic medication is considered the primary treatment of FOG, we determined the most reliable protocol to detect medication effects on FOG-severity at two timepoints. The same methodology was applied as in Supplementary Material 6, but with %TF OFF-ON as outcome instead of %TF. Second, we compared sensitivity and test-retest reliability of several tasks and task combinations to detect changes in FOG-severity in OFF compared to ON. One side note is that medication effects on %TF are in practice more often determined at one timepoint only. Nonetheless, as the effect of chance of FOG-provocation still plays a role, it is useful to evaluate reliability over time.

*Results*

The 360° turns ST and Hotspot door did not show significant differences between OFF and ON (Supplementary Material 3), and may therefore be less sensitive for medication effects. Test-retest reliability for %TF OFF-ON was not very good (see Supplementary Table 10). Despite good average ICCs for most tasks, MDC index values of >1 indicate a large test-retest error. We further found that the best performing tasks on reliability for FOG-severity were not necessarily those that could most reliability detect medication effects on FOG-severity. The lowest MDC index for %TF OFF-ON was 1.27, and was found for a combination of: TUG DT, Hotspot door, Personalized hotspot. However, this protocol performed poorly on reliability of %TF, with MDC index values of 1.35, 1.19 and 2.25 in OFF+ON, OFF and ON. In contrast, the optimal protocol for %TF (TUG ST, 360° turns ST, 360° turns DT and Hotspot door), had an MDC index for %TF OFF-ON of 1.95, but MDC index values for %TF of 0.48, 0.53 and 0.90 in OFF+ON, OFF and ON.

**Supplementary Table 10. Sensitivity and test-retest reliability for medication effects on FOG-severity, per condition.**

| | 4MW | TUG ST | TUG DT | 360° turns ST | 360° turns DT | Hotspot door | Personal. hotspot | Optimal protocol for %TF[#] | Optimal protocol for %TF OFF-ON^ | All tasks combined |
|---|---|---|---|---|---|---|---|---|---|---|
| %TF OFF-ON[a], mean (SD) | 3.75 (12.9) | 6.39 (13.2) | 10.1 (20.1) | 9.40 (30.3) | 13.6 (32.5) | 6.65 (20.0) | 12.7 (16.1) | 8.92 (19.1) | 10.1 (14.3) | 9.02 (14.6) |
| Δ%TF OFF-ON[b], mean (SD) | 5.17 (19.1) | 2.38 (12.7) | -4.66 (15.0) | -1.75 (24.7) | -3.72 (22.4) | -2.88 (19.4) | -2.22 (14.4) | -1.91 (11.3) | -2.87 (7.82) | -1.12 (8.99) |
| ICC %TF OFF-ON[b], mean (95% CI) | -0.14 (-0.49; 0.25) | 0.49 (0.14; 0.74) | 0.70 (0.43; 0.85) | 0.68 (0.39; 0.85) | 0.72 (0.45; 0.87) | 0.53 (0.17; 0.77) | 0.73 (0.47; 0.87) | 0.83 (0.64; 0.92) | 0.88 (0.73; 0.95) | 0.84 (0.66; 0.93) |
| MDC %TF OFF-ON[b] | 37.5% | 24.8% | 29.5% | 48.5% | 43.9% | 37.9% | 28.3% | 22.1% | 15.3% | 17.6% |
| MDC index %TF OFF-ON[b] | 11.2 | 4.17 | 2.38 | 4.03 | 2.49 | 3.58 | 2.12 | 1.95 | 1.27 | 1.65 |

4MW, Four meter walk; TUG, Timed-Up and Go; ST, single task; DT, dual task; FOG, freezing of gait; %N, percentage of people relative to the total; %TF, percentage time frozen; SD, standard deviation; OFF, >12 h withdrawal of dopaminergic medication; ON, 1 h after intake of dopaminergic medication; OFF-ON, medication effect (value in OFF minus ON state); Δ%TF, $\%TF_{T2} - \%TF_{T1}$; ICC, intra-class correlation coefficient; CI, confidence interval; MDC, minimal detectable change; MDC index, MDC/mean %TF at T1. Outcomes are categorized as [a] single timepoint (T1, N=63), [b] comparison between two timepoints (T1 vs. T2, N=26). [#] optimal protocol for FOG-severity (lowest MDC index for %TF)= TUG ST+ 360° turns ST+ 360° turns DT+ Hotspot door. ^ optimal protocol for medication effects (lowest MDC index for %TF OFF-ON)= TUG DT+ Hotspot door+ Personalized hotspot.

## REFERENCES


1. Zoetewei D, Herman T, Brozgol M, et al. Protocol for the DeFOG trial: A randomized controlled trial on the effects of smartphone-based, on-demand cueing for freezing of gait in Parkinson's disease. *Contemp Clin Trials Commun* 2021; 24: 100817.
2. Zoetewei D, Herman T, Ginis P, et al. On-demand cueing for freezing of gait in Parkinson's disease: a randomized controlled trial. *Mov Disord* 2024; 39: 876–886.
3. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Front Psychol* 2013; 4: 863.