Neuron, Volume *70*

Supplemental Information

Triangulating the Neural, Psychological, and Economic Bases of Guilt Aversion

Luke J. Chang, Alec Smith, Martin Dufwenberg, and Alan G. Sanfey

Supplemental Information Inventory

Figures

Figure S1 related to Equation 1. Simulation of Guilt-Aversion Model
The guilt-aversion model makes two behavioral predictions depending on Θ_{12}. If $\Theta_{12} < 1$, then the optimal choice (S_2) that maximizes the utility function (U_2) is to keep all of the money. If $\Theta_{12} > 1$, then the optimal choice which maximizes U_2 is to match expectations and return the amount that player 2 believes that player 1 expects them to return. Here we plot the behavioral predictions for Player 2's choice if Player 1 invests \$10 (which become \$40) and Player 2 believes that Player 1 expects them to return \$20 for varying Θ_{12}s (e.g., 0.5 or 1.5).

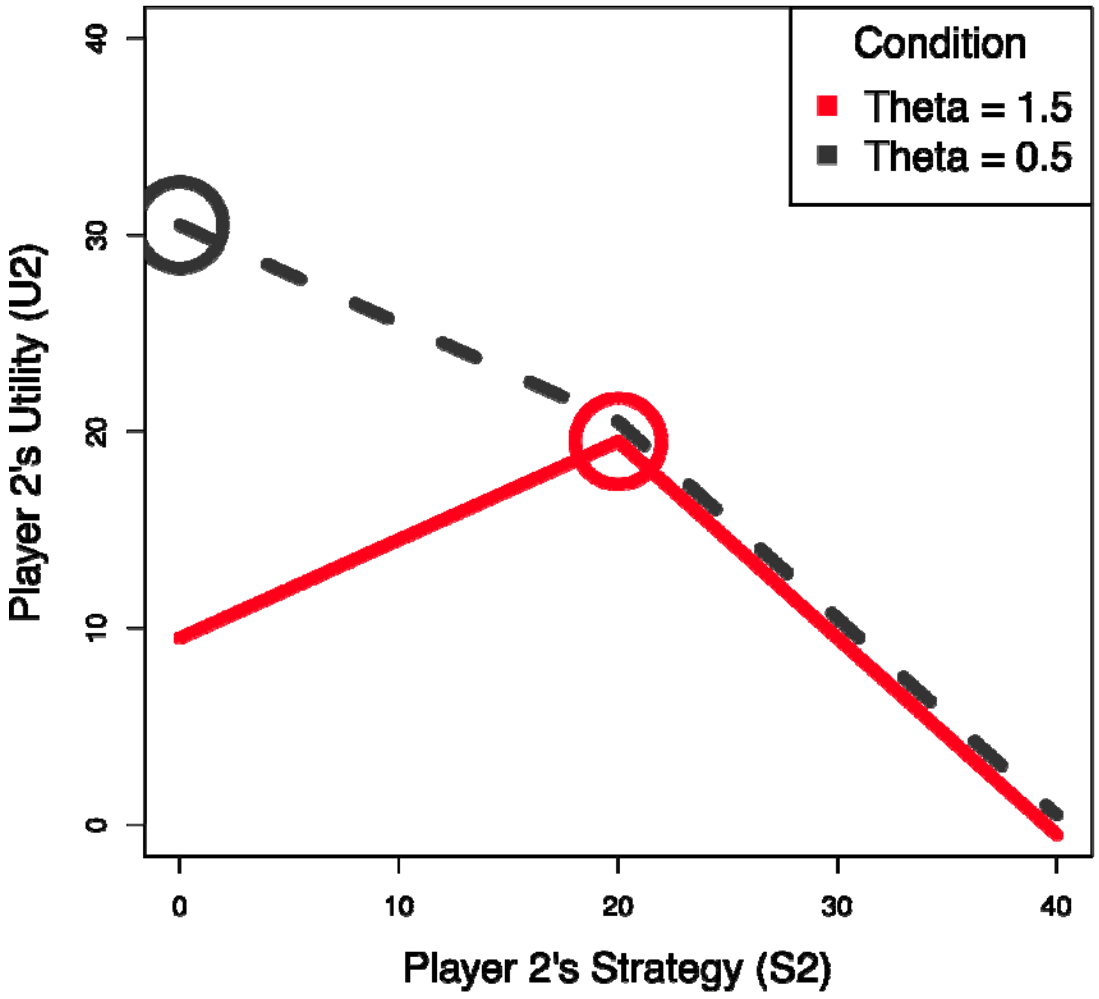


Figure S2 related to Figure 4. Relationship between SMA and Guilt Sensitivity. Participant's best linear unbiased predictors (BLUPs) from the counterfactual guilt analysis predict the average parameter estimate of voxels in the SMA ROI using robust regression.

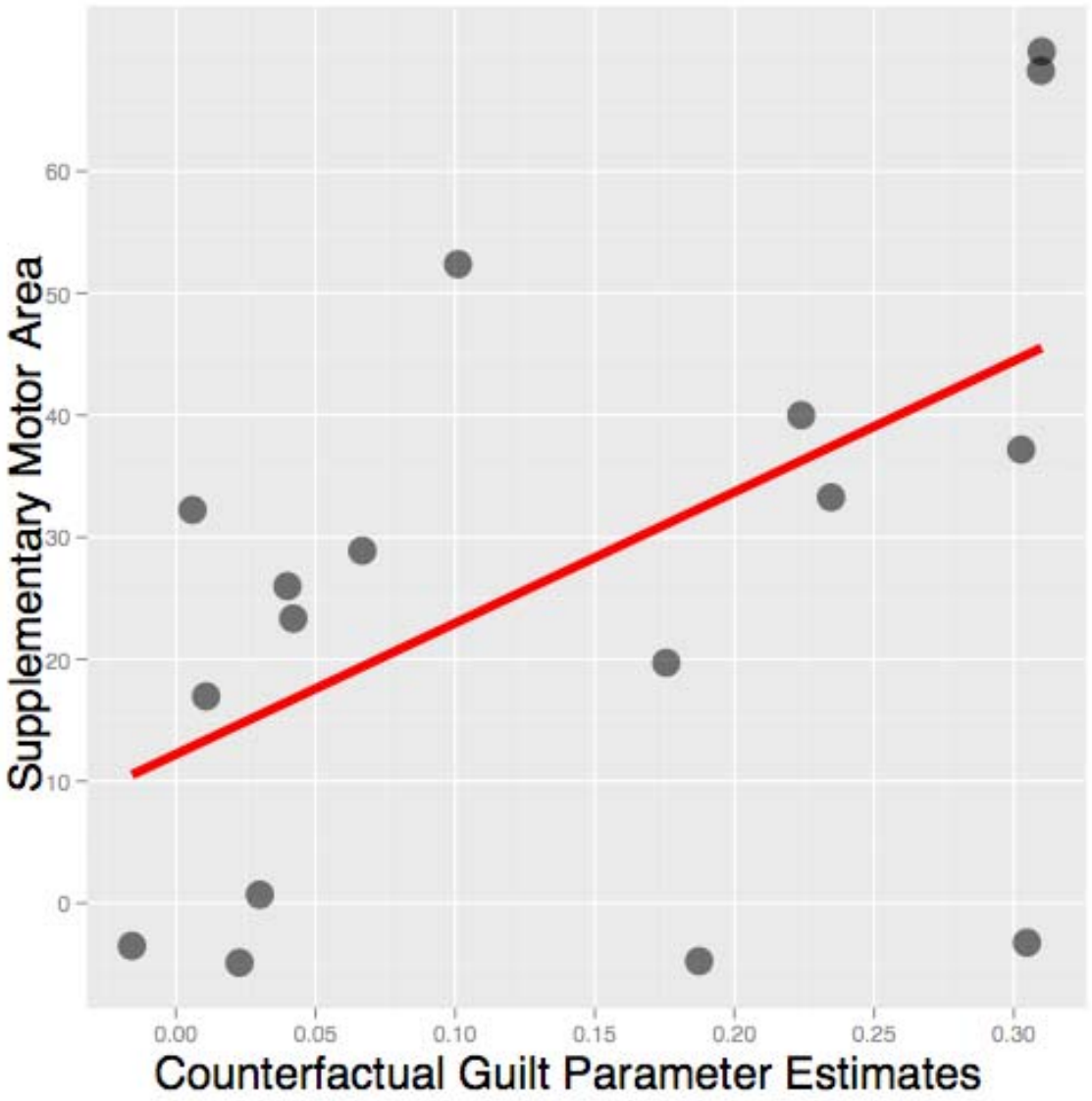
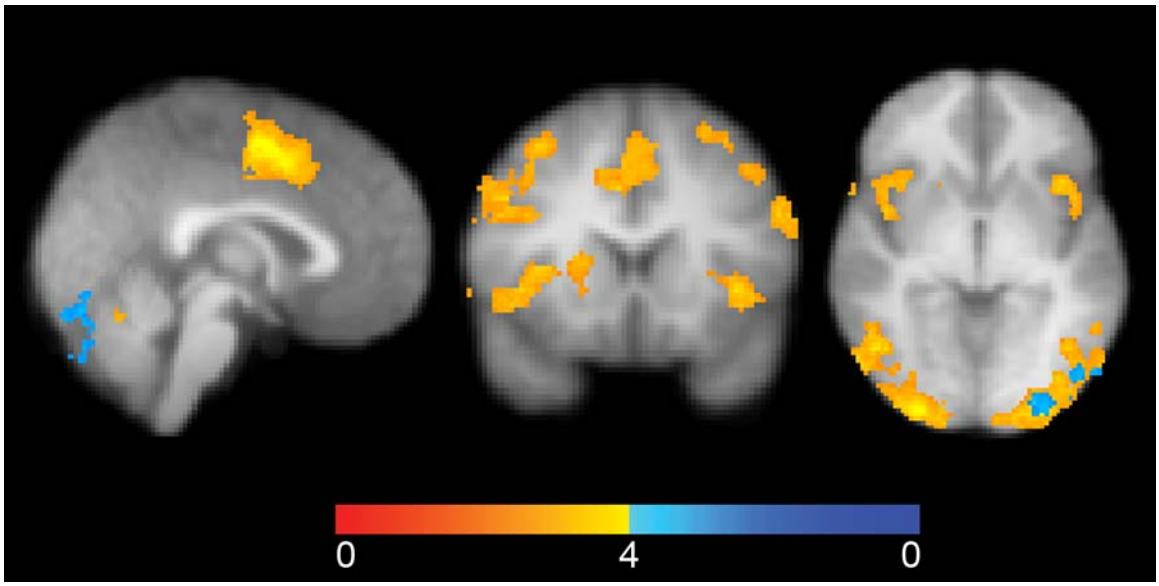


Figure S3 related to Figure 5. Guilt-Aversion Controlling for Player 2's Choice. This figure depicts activity associated with matching expectations (orange) and returning less than participants believed their partner expected them to return (blue). Images are displayed in radiological orientation (left=right) and are thresholded using whole brain cluster correction, $Z > 2.3$, $p < 0.05$. Color maps reflect Z values between 0 and 4.



Tables

Table S1 (related to figure 3). The Effect of Beliefs on Trustee Behavior. This table illustrates the results of a mixed effects regression analysis, in which Trustee's 2nd order beliefs (E2E1S2) significantly predict the amount of money that they chose to return to the Investor (S2) controlling for the size of the initial investment (Offer Amount). Participants were treated as a random effect with varying intercepts (s^2=0.004, SD=0.06). All variables were normalized to [0,1]. There was a correlation of 0.34 between the two fixed effects predictors. These results indicate that there is a significant effect of expectations on behavior controlling for the subgame.

Predictor	Parameter Estimate	SE	t - Value	p - Value
Intercept	0.09	0.03	3.54	< .001
Offer Amount	0.17	0.03	6.82	< .001
E2E1S2	0.46	0.04	10.83	< .001

Table S2 (related to Figure 4). Brain activations for Matching Compared to Returning Less then Expectations Contrast. This table reflects the contrast matching expectations (i.e. 2nd order beliefs) compared to returning less than expectations and shows the local maxima of clusters surviving cluster correction $Z > 2.3$, $p < 0.05$ in MNI space. Cortical and subcortical regions were identified using the Harvard-Oxford Probabilistic Anatomical Atlas, while the cerebellar regions were identified using a probabilistic cerebellar atlas (Diedrichsen et al., 2009). Abbreviations: DMPFC=dorsomedial prefrontal cortex, DLPFC=dorsolateral prefrontal cortex, TPJ = temporal-parietal junction, SMA = supplementary motor area, OFC = orbitofrontal cortex, ACC = anterior cingulate cortex.

	Hemisphere	Region	BA	Z Value	X	Y	Z
Less > Equal							
	L	dmPFC (Frontal Pole)	10	3.61	-12	64	24
	L	Nucleus Accumbens	25	2.99	-8	8	-12
	L	Paracingulate	10	3.41	-6	54	6
	L	Superior Frontal Gyrus (DMPFC)	9	3.76	-12	46	50
	L	Medial OFC	11	3.16	-8	26	-16
	R	Caudate	25	3.11	14	20	4
	R	Medial OFC	11	3.1	4	28	-16
	R	Nucleus Accumbens	25	2.92	6	14	-4
	R	Rostral ACC	10	3.39	14	46	0
	R	Sub Genual ACC	25	3.28	0	14	-14
	R	Superior Frontal Gyrus (DMPFC)	10	3.51	4	56	26
Equal > Less							
	L	ACC	24	3.51	-4	18	34
	L	Cerebellum (Left Crus I)	19	3.65	-42	-74	-26
	L	Middle Frontal Gyrus (DLFPC)	45	3.49	-48	30	26
	L	Fusiform	19	3.53	-44	-70	-20
	L	Lateral Occipital Cortex	37	3.49	-56	-64	6
	L	Posterior Cingulate Cortex	NA	3.4	-12	-28	42
	L	Postcentral gyrus	3	4.27	-34	-30	58
	L	Precentral Gyrus	6	4.17	-34	-4	48
	L	Precentral Gyrus	43	3.96	-58	2	28
	L	SMA	NA	3.75	-4	-2	48
	L	Superior Parietal Lobule	40	4.22	-34	-40	50
	L	Supramarginal Gyrus (TPJ)	2	3.6	-54	-36	36
	R	ACC	24	3.32	6	16	36
	R	Cerebellum (Right Crus I)	37	3.9	36	-58	-30
	R	Cerebellum (Right VI)	NA	3.92	30	-46	-38
	R	Cerebellum (Vermis VI)	NA	3.75	6	-66	-22
	R	Middle Frontal Gyrus (DLPFC)	46	3.5	36	42	28
	R	Inferior Temporal gyrus	37	4.12	52	-50	-22
	R	Insula	48	3.42	42	8	0
	R	Lateral Occiptal Cortex, Inferior division	19	4.18	50	-72	-8
	R	Lateral Occiptal Cortex, Superior division	7	3.82	26	-62	42
	R	Occipital Pole	18	3.63	30	-94	-10
	R	SMA	NA	3.64	2	2	50
	R	Superior Parietal Lobule	40	3.88	36	-46	46
	R	Supramarginal Gyrus (TPJ)	2	3.48	50	-28	36

Table S3 (related to Figure 5). Brain activations for Parametric Contrast of Matching Compared to Returning Less then Expectations. This table reflects the parametric contrast matching expectations (i.e. 2nd order beliefs) compared to returning less than expectations (i.e. 10%, 20%, +30%) and shows the local maxima of clusters surviving cluster correction $Z > 2.3$, $p < 0.05$ in MNI space. Cortical and subcortical regions were identified using the Harvard-Oxford Probabilistic Anatomical Atlas, while the cerebellar regions were identified using a probabilistic cerebellar atlas (Diedrichsen et al., 2009). Abbreviations: DMPFC=dorsomedial prefrontal cortex, DLPFC=dorsolateral prefrontal cortex, TPJ = temporal-parietal junction, SMA = supplementary motor area, OFC = orbitofrontal cortex, ACC = anterior cingulate cortex.

	Hemisphere	Region	BA	Z Value	X	Y	Z
Less > Equal							
	L	Caudate	25	3.08	-10	18	4
	L	Lateral OFC Cortex	47	4.06	-36	32	-18
	L	Middle Frontal Gyrus	9	3.52	-28	24	42
	L	Nucleus Accumbens	25	3.3	-6	10	-8
	L	Paracingulate Gyrus	9	3.7	-6	54	6
	L	Rostral ACC	32	3.18	-10	40	20
	L	Sub Genual ACC	25	3.54	-4	14	-16
	L	Superior Frontal Gyrus (DMPFC)	9	4.11	-12	46	50
	L	Superior Frontal Gyrus (DMPFC)	8	3.69	-20	26	58
	L	Superior Frontal Gyrus (DMPFC)	10	3.92	-12	64	24
	L	Temporal Pole	38	3.51	-42	20	-30
	R	ACC	11	3.38	0	30	-6
	R	Caudate	25	3.19	12	20	4
	R	Lateral OFC Cortex	11	3.25	22	36	-16
	R	Medial OFC Cortex	11	3.51	4	44	-20
	R	Nucleus Accumbens	25	3.08	8	16	-4
	R	Paracingulate Gyrus	10	4.05	14	46	0
	R	Posterior Insula	48	3.06	40	0	-14
	R	Rostral ACC	11	3.21	8	36	2
	R	Superior Frontal Gyrus (DMPFC)	10	3.94	2	56	32
	R	Temporal Pole	28	3.35	26	10	-26
	R	Temporal Pole	38	3.24	42	20	-20
Equal > Less							
	L	Angular Gyrus	19	3.45	-44	-56	44
	L	Cerebellum (Crus I)	NA	3.61	-40	-74	-26
	L	Dorsal ACC	32	4.02	-8	14	38
	L	Fusiform	19	3.57	-44	-70	-20
	L	Lateral Occipital Cortex, Inferior Division	37	3.91	-58	-64	8
	L	Lateral Occiptal Cortex, Inferior Division	19	3.81	-52	-74	-14
	L	Middle Frontal Gyrus (DLPFC)	45	3.88	-44	28	24
	L	Middle Frontal Gyrus (DLPFC)	46	3.08	-32	40	24
	L	Postcentral Gyrus	40	4.74	-38	-34	40
	L	Precentral Gyrus	6	4.48	-34	-6	48
	L	Precentral Gyrus	4	4.43	-58	0	30
	L	SMA	NA	3.85	-4	-2	48
	L	Superior Parietal Lobule	40	4.05	-34	-52	58
	L	Supramarginal Gyrus	2	3.88	-58	-28	44
	L	Supramarginal Gyrus (TPJ)	40	4.01	-46	-34	38
	R	Central Opercular Cortex	48	3.58	48	-2	10

R	Cerebellum (I-IV)	NA	3.71	4	-52	-16
R	Cerebellum (Right Crus I)	NA	4.1	36	-58	-30
R	Cerebellum (Vermis VI)	NA	4.22	6	-64	-22
R	Cerebellum (VI)	NA	4.17	30	-46	-38
R	Cerebellum (X)	NA	3.62	30	-36	-44
R	DLPFC	46	3.64	36	42	28
R	Dorsal ACC	24	3.53	6	16	36
R	Fusiform	37	3.68	42	-48	-20
R	Inferior Temporal Gyrus	37	4.09	48	-60	-12
R	Insula	48	3.73	46	14	-2
R	Lateral Occipital Cortex, Inferior Division	19	4.21	50	-72	-8
R	Lateral Occipital Cortex, Superior Division	7	4.12	26	-62	42
R	Occipital Pole	18	3.97	30	-92	-8
R	Precentral Gyrus	6	4.07	46	-4	56
R	Precuneous	7	3.81	6	-70	48
R	SMA	6	3.94	6	-6	64
R	Supramarginal gyrus	40	4.19	44	-40	48

Table S4 (related to figure S3). Brain Activations for Guilt-Aversion Controlling for Player 2's Choice. This table reflects the activity associated with matching expectations (i.e., (E2E1S2-S2)=0) and returning less than participants believed their partner expected them to return (i.e., $(E2E1S2-S2)^{+}$) controlling for Player 2's choice (i.e., S2) and shows the local maxima of clusters surviving cluster correction $Z > 2.3$, $p < 0.05$ in MNI space. Cortical and subcortical regions were identified using the Harvard-Oxford Probabilistic Anatomical Atlas, while the cerebellar regions were identified using a probabilistic cerebellar atlas (Diedrichsen et al., 2009). Abbreviations: DLPFC=dorsolateral prefrontal cortex, TPJ = temporal-parietal junction, SMA = supplementary motor area, OFC = orbitofrontal cortex, ACC = anterior cingulate cortex, STS = superior temporal sulcus.

	Hemisphere	Region	BA	Z Value	X	Y	Z
Return Less than Expectations							
	L	Cerebellum (Crus I)	NA	3.47	-38	-76	-28
	L	Cerebellum (Crus II)	NA	3.32	-6	-80	-30
	L	Lateral Occipital Cortex, Inferior Division	18	3.53	-32	-88	-16
	L	Lateral Occipital Cortex, Superior Division	7	3.32	-22	-66	40
	L	Middle Temporal Gyrus	20	3.23	-60	-32	-18
	R	Superior Parietal Lobule	7	3.38	32	-54	54
Match Expectations							
	L	DLPFC (Middle Frontal Gyrus)	46	3.42	-38	36	28
	L	Insula (Anterior)	48	3.69	-38	14	-2
	L	Insula (Posterior)	48	3.9	-42	-2	6
	L	Lateral Occipital Cortex, Inferior Division	19	4.35	-44	-76	-16
	L	Lateral Occipital Cortex, Superior Division	19	4.41	-26	-70	26
	L	Postcentral Gyrus	4	4.85	-48	-18	52
	L	Precentral Gyrus	6	4.02	-56	6	32
	L	SMA	NA	4.32	-2	-2	54
	L	Supramarginal Gyurs (TPJ)	2	4.11	-48	-30	36
	R	ACC	24	3.39	6	16	36
	R	Caudate	NA	3	18	14	8
	R	Cerebellum (V)	37	3.53	14	-52	-22
	R	Cerebellum (Crus I)	19	3.51	38	-76	-24
	R	Cerebellum (VI)	NA	3.79	16	-58	-28
	R	DLPFC (middle frontal gyrus)	46	3.88	34	36	28
	R	Insula (Anterior)	47	3.41	42	18	-6
	R	Insula (Middle)	48	3.27	44	4	-2
	R	Lateral Occipital Cortex, Inferior Division	19	4.45	34	-86	-8
	R	Lateral Occipital Cortex, Inferior Division	37	4.32	52	-64	-8
	R	Lateral Occipital Cortex, Superior Division	7	4.12	34	-60	44
	R	Parietal Operculum Cortex (TPJ)	48	3.23	56	-22	16
	R	Precentral Gyrus	6	4.27	60	10	32
	R	Precentral Gyrus	6	3.92	28	-8	48
	R	Supramarginal Gyrus, Posterior Division	40	4.09	42	-44	46

R	Supramarginal Gyrus, Posterior Division (TPJ)	22	3.93	66	-44	18
R	Temporal Occipital Fusiform Cortex	37	3.86	38	-48	-24

Supplemental Experimental Procedures

Guilt-Aversion Controlling for Player 2's Behavior: As a consequence of our design, participants make systematically less money in trials in which they match expectations compared to trials in which they return less than they believe the other player expected them to return. Presumably, participants choose to return more money to Player 1 because they are more motivated by minimizing guilt aversion then maximizing financial payoff. However, to rule out the possibility that the insula is simply tracking forgone financial payoffs rather than guilt-aversion, we ran an analysis which allowed us to examine the effect of matching expectations (i.e., guilt-aversion=0 in eq(1)), while controlling for the amount of money that they choose to return (i.e., their forgone financial payoff or S2). This model included the following regressors:

1) Return phase
2) Guilt (i.e., E2E1S2-S2)+
3) Match Trials (i.e., Guilt = 0)
4) Player 2's Choice (i.e., S2)
5) Over Match Trials (i.e., E2E1S2-S2)-
6) Face phase
7) Prediction phase
8) Investment phase
9) Belief elicitation phase
10) Summary phase
11) Handed-down-belief phase
12) Missed trials
13-24) Temporal derivatives of regressors 1 – 12
25-31) Estimated head movement parameters (n=6)

We report the results for the independent variance associated with matching expectations (regressor 3) and linear deviations of returning less money than participants believed Player 1 expected (regressor 2), while controlling for all of the other regressors in Figure S3 and Table S4. We were forced to exclude 8/66 runs due to a lack of variability in either regressor 2 or 3.

Supplemental References

Diedrichsen, J., Balsters, J.H., Flavell, J., Cussans, E., and Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. Neuroimage *46*, 39-46.