## Supplementary material for “Rethinking the cognitive mechanisms underlying pantomime of tool use: evidence from Alzheimer’s disease and semantic dementia”

To validate our method, 71 control participants took part in the four different tasks described above. Then, we matched our 30 AD patients with 30 new control subjects who were not part of the 71 abovementioned participants. The distribution of scores of PTU is displayed in **Fig.1**. An accuracy-based scoring system (Jarry et al., 2013) gave strong ceiling effect and the scores were obviously not normally distributed, whether for the 71 control participants or for the 30 control participants (*W* = .92, *p*< .001 and *W* = .90, *p*< .05, respectively; **Fig.1a**). After transformation of the data, scores were normally distributed in both groups (*W* = .98, *p* = .30 and *W* = .97, *p* = .49; **Fig.1b**). Moreover, a Kolmogorov-Smirnov test showed that these two samples were not different from each other, suggesting that they came from the same distribution (*D* = .16, *p* = .65). This data transformation showed similar results concerning MPS, RTM and FM tasks.



**Figure 1.** Frequency distribution of scores for the Pantomime of tool use task (PTU) obtained using two different scoring systems. a) Accuracy-based scoring system (see Jarry et al., 2013). b) Time-based scoring system (Lesourd et al., 2015).



In the PTU task, each item was scored on a 10 point-scale relative to the completion time of the item so the maximum score was 100. If a production (i.e., BPO or spatiotemporal) error was made, only 1 point was given and if a conception (i.e., content or perplexity) error was made, no point was allocated. By summing the 10 scores obtained for each item, the maximum score was 100.

****

**Fig. 2** Frequency distribution of scores for Mechanical Problem-Solving task in the no choice condition (MPS-NC) obtained using two different scoring systems. a) Accuracy-based scoring (see Jarry et al., 2013: the scores were not normally distributed for the 71 control participants and for the 30 control participants (*W* = .37, *p* < .001 and *W* = .56, *p* < .001, respectively). b) Time-based scoring system: After transformation of the data, the distributions of scores were normally distributed for the 71 control participants and for the 30 control participants (*W* = .98, *p* = .42 and *W* = .98, *p* = .79, respectively). A Kolmogorov-Smirnov test showed that these two samples were not different from each other, suggesting that they came from the same distribution (*D* = .10, *p* = .99).



In the MPS task, each box was rated on a 10 point-scale relative to the time of completion of the item. For completion time beyond C95, 3 points were accorded. Only 2 points were attributed if a critical action was performed but the target was still in the box at the end of the 3 minutes (e.g., Box 1: The target was pushed down with one rod from the “chimney” but was still in the box). Only 1 point was given if the participant touched the target with a rod but no critical action was performed during the 3 minutes (e.g., Box 1: the participant touched the red wooden cube with rod A but did not push down the target from the “chimney”). In the other cases, no point was allowed. By summing the three scores obtained for each box, the maximum score was 30.



**Fig. 3** Frequency distribution of scores for Recognition of Tool Manipulation (RTM) obtained using two different scoring systems. a) Accuracy-based scoring (see Jarry et al., 2013: the scores were not normally distributed for the 71 control participants and for the 30 control participants (*W* = .69, *p* < .001 and *W* = .84, *p* < .001, respectively). b) Time-based scoring system: After transformation of the data, the distributions of scores were normally distributed for the 71 control participants and for the 30 control participants (*W* = .97, *p* = .13 and *W* = .97, *p* = .65, respectively). A Kolmogorov-Smirnov test showed that these two samples were not different from each other, suggesting that they came from the same distribution (*D* = .22, *p* = .25).



For both FM and RTM tasks, each item was rated on a 10 point-scale relative to the response time of the participant for the item. For response time beyond C95, only 1 point was attributed if the participant gave the correct answer. If the participant gave a wrong answer, whatever the response time, no point was allowed. By summing the ten scores obtained for each item, the maximum score was 100.



**Fig. 4** Frequency distribution of scores for Functional Matching (FM) obtained using two different scoring systems. a) Accuracy-based scoring (see Jarry et al., 2013: the scores were not normally distributed for the 71 control participants and for the 30 control participants (*W* = .77, *p* < .001 and *W* = .81, *p* < .001, respectively). b) Time-based scoring system: After transformation of the data, the distributions of scores were normally distributed for the 71 control participants and for the 30 control participants (*W* = .97, *p* = .13 and *W* = .94, *p* = .09, respectively). A Kolmogorov-Smirnov test showed that these two samples were not different from each other, suggesting that they came from the same distribution (*D* = .17, *p* = .58).



