



ACADEMIC
PRESS

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

NeuroImage

NeuroImage 20 (2003) 572–577

www.elsevier.com/locate/ynimg

Rapid Communication

Equivalence of cognitive processes in brain imaging and behavioral studies: evidence from task switching

Iring Koch,^{a,*} Hannes Ruge,^b Marcel Brass,^b Orit Rubin,^c Nachshon Meiran,^c
and Wolfgang Prinz^a

^a Max Planck Institute for Psychological Research, Munich, Germany

^b Max Planck Institute of Cognitive Neuroscience, Leipzig, Germany

^c Ben-Gurion University of Negev, Beer-Sheva, Israel

Received 23 December 2002; revised 5 March 2003; accepted 3 April 2003

Abstract

A growing number of studies on the higher-order cognitive functions of the human brain use brain-imaging techniques, such as functional magnetic resonance imaging (fMRI). For the validity and generality of fMRI results, it is important that the relevant cognitive processes are equivalent to those functioning in typical settings used in behavioral research. This equivalence could be, for example, endangered by different spatial frames of reference when lying in the scanner. In the present study, we tested whether the cognitive processes, as reflected in behavioral data in brain-imaging settings, are indeed functionally equivalent to those reflected in “purely” behavioral settings. To this end, we used a task-switching paradigm with a spatial component, increasing the likelihood to find effects of experimental setting. We compared the data of three different groups that only differed in testing environments (real, operating fMRI vs simulated fMRI vs standard behavioral with upright position of participants) but used otherwise strictly equivalent experimental conditions. Of importance for our validation purposes, unlike previous studies, we included a group with a behavioral setting, and we tested whether we would replicate a nontrivial, complex three-way interaction across all three groups. We replicated the predicted complex data pattern in all groups, suggesting functional equivalence of the underlying cognitive processes. We also found strongly increased reaction time (RT) levels in the two fMRI groups. We attribute this increase to unspecific distracting factors affecting late motor processes and discuss potential methodological implications of this increased baseline RT in the scanner.

© 2003 Elsevier Science (USA). All rights reserved.

A growing number of studies on the cognitive functions of the human brain use brain-imaging techniques. Two major imaging techniques are position emission tomography (PET) and functional magnetic resonance imaging (fMRI). Most notably during the past years, the enhanced spatial and temporal resolution of fMRI has led to an increasingly better understanding of the neural basis of cognitive higher-order functions (see, e.g., D’Esposito et al., 1999).

A major merit of using brain-imaging techniques is that imaging data provide a potentially very important measure

correlated with cognitive functions in addition to the more traditional measures of cognitive psychology, such as reaction time (RT) and error rate. Integrating conclusions from these complementary measures may thus have the power to give a strong synergy effect in cognitive neuroscience.

It is clear though that the validity of the conclusions derived from brain-imaging studies critically rests on the assumption that the use of brain-imaging techniques does not alter the cognitive functions the experimenter sets out to explore. In the present study, we aimed at testing whether the cognitive processes in imaging experiments are indeed functionally equivalent to those in behavioral experiments.

There are at least two reasons why cognitive processes might be altered in the fMRI scanner relative to standard experimental settings. First, it might be that adaptations of the experimental design in order to comply with the needs

* Corresponding author. Max-Planck-Institut für Psychologische Forschung, Amalienstrasse 33, 80799 München, Germany. Fax: +49-0-89-38602-190.

E-mail address: koch@psy.mpg.de (I. Koch).

of newly developed techniques, such as event-related fMRI, affect the very nature of the cognitive functions one is interested in. For instance, using very long intertrial intervals deviates from the typically much shorter intervals in cognitive psychology studies.

Second, it might be that the very unfamiliar and uncomfortable environment inside the fMRI scanner has relevant effects on cognitive processes. For instance, inside the scanner, the participant is in a supine position, it is quite noisy, and task stimuli are presented by means of special lenses or a mirror. In contrast, conditions of typical behavioral experiments are that participants sit in an upright position in front of the screen in a noise-free lab. Consequently, the spatial frame of reference of participants may be very different between these settings. This aspect was especially emphasized in the present study, in which participants were required to judge the vertical position of a target stimulus within a 2×2 grid. When sitting, this judgment was between UP and DOWN; however, when lying down, this judgment was, properly speaking, between “in front” and “behind”. Hence, in particular for tasks in which spatial information is relevant, it might be that these differences in the experimental settings also change some aspects of cognitive processing.

The goal of the present study was to test whether cognitive processes are indeed functionally equivalent in brain-imaging settings and in behavioral settings. We tested this important methodological question by using a task-switching paradigm. The task-switching paradigm has been used in a number of recent imaging studies exploring cognitive control processes (e.g., Brass and von Cramon, 2002; Dove et al., 2000; Kimberg et al., 2000; Sohn et al., 2000). In the task-switching paradigm, performance typically declines in a task switch as compared to a task repetition. These “shift costs” may be related to cognitive control processes making sure that the correct task is performed (e.g., Allport et al., 1994; Meiran, 1996; Rogers and Monsell, 1995). Given the current prominence of research on cognitive control in cognitive neuroscience (see Miller and Cohen, 2001), it is of particular importance to test the functional equivalence of cognitive control processes inside and outside the scanner.

In order to investigate this functional equivalence in task switching, we tested three groups of participants. The tasks and experimental design were identical in all groups. In a first group, we tested participants in a lying position on a stretcher (“simulated MRI” group, for details see Methods). The second group was tested under real fMRI conditions inside the operating scanner (“real MRI”). Of importance, we also tested a third group with the standard conditions of a behavioral experiment (“standard behavioral”).

In all three groups, we used a pair of spatial judgment tasks introduced by Meiran (1996). A spatial target stimulus is presented in one of four quadrants on the screen. The task is either to judge the vertical stimulus position (i.e., whether it is “up” or “down”, ignoring the horizontal stimulus dimension) or to judge the horizontal position (left vs right,

ignoring the vertical dimension). Participants indicate their judgment by pressing either a lower left key (for down or left) or an upper right key (for up or right). Note that these tasks involve strong spatial components, thus increasing the likelihood of finding differences across settings.

Because the S–R mappings of both tasks are completely overlapping, an external task cue indicated the current task in an otherwise random task sequence. In this paradigm, task preparation processes can be explored by varying the cue-target interval. The typical result is that performance improves with longer cuing intervals. Of importance, shift costs are also reduced, indicating a process of advance reconfiguration of task set (e.g., Meiran, 1996; Rogers and Monsell, 1995). According to Meiran’s (2000) model, one aspect of task-set reconfiguration consists of changing attentional weights of the currently relevant and irrelevant stimulus dimensions. This stimulus-set reconfiguration should not be necessary when the stimuli are univalent (i.e., contain only one dimension), so both shift costs and cuing effects should be reduced as compared with bivalent stimuli. Note that we derived from this model a complex, nontrivial three-way interaction. We tested whether we would replicate this complex interaction across experimental settings, thus providing a powerful test of the functional equivalence of cognitive control processes inside and outside the scanner.

In summary, in order to test the functional equivalence of cognitive processes in brain-imaging studies and behavioral studies, we (1) used a paradigm with a strong spatial component, thus rendering effects of experimental setting likely; (2) tested the replicability of a complex, theoretically derived three-way interaction; and (3) compared this interaction in three different groups, including one with standard behavioral setting.

Methods

Participants

We tested three groups of participants. Thirty-six participants were tested at the Max Planck Institute for Psychological Research in Munich and were randomly assigned to either the standard behavioral group (10 female, 8 male, age 19–29) or the simulated MRI group (13 female, 5 male, age 18–28). The real MRI group of 16 participants (8 female, 8 male, age 22–32) was tested inside the operating scanner at the Max Planck Institute of Cognitive Neuroscience in Leipzig.

Task, apparatus, and procedure

Participants performed two spatial judgment tasks. In one task, they were required to decide whether a square (1.0-cm side length), presented in a two by two grid (6.0-cm side length), was in the upper or lower half of the grid, and

Table 1
Mean RT as a function of stimulus valence, trial type, cuing interval, and group

Group	Univalent stimuli			Bivalent stimuli		
	Trial type			Trial type		
	Shift	Repeat	Shift cost	Shift	Repeat	Shift cost
Standard behavioral						
Cuing interval = 100 ms	415	388	27	566	499	67
Cuing interval = 2000 ms	409	399	10	482	461	21
Simulated MRI						
Cuing interval = 100 ms	474	446	28	647	553	94
Cuing interval = 2000 ms	463	455	8	540	524	16
Real MRI						
Cuing interval = 100 ms	558	536	22	720	643	77
Cuing interval = 2000 ms	535	525	10	592	578	14

in the other task, they decided whether the square was on the left or right side of the grid. The task in a given trial was indicated by two arrows (0.6 cm wide) presented either above or below the grid, or on the left or right side of the grid, indicating the up–down task or the left–right task, respectively. In both tasks, participants responded with their index fingers of the left and right hand. Half of the participants used either a lower left key and an upper right key on an external keyboard, whereas the other half of the participants used an upper left key and a lower right key.

Participants were instructed about the tasks and that they would perform these tasks in two different stimulus-valence conditions. They received demonstration trials in order to become familiar with the experimental conditions. In the bivalent stimulus block, the stimuli were presented in one out of four spatial positions on the screen. In the univalent block, the stimuli were presented either in the upper or lower part of the screen, superimposed on the vertical dividing line, or in the left or right part of the screen, superimposed on the horizontal dividing line. The experiment consisted of two experimental blocks of 128 trials each. In each block, 16 empty trials were randomly inserted. The order of blocks was counterbalanced across participants. Before each of the experimental blocks, a short practice phase of 60 trials was carried out in order to familiarize participants with the stimulus layout.

A trial started with the presentation of the grid for 300 ms. The task cues were then presented, and after a randomly chosen period of 100- or 2000-ms cue-target interval, the target appeared. Participants responded by pressing one out of two response keys. After that, an empty response-grid interval was presented for a variable duration, so that the intertrial interval was 6000 ms. We introduced a jitter in the onsets of events for the purpose of the real event-related fMRI measurement, but that was identical for the groups. Exactly the same experimental software and experimental definitions were used for all three groups, except for the differences described below.

In the standard behavioral group, participants sat in a silent room in an upright position in front of a VGA monitor

connected to an IBM compatible PC. In the simulated MRI group, participants lay down on a stretcher. They viewed the stimulus display positioned behind their head via an individually adjustable mirror. They heard typical MRI noise via a headphone. Participants were told to adjust the volume so that it was uncomfortable but still bearable. While these two groups were tested at the Max Planck Institute for Psychological Research in Munich, the third, real fMRI group was tested inside an operating scanner at the Max Planck Institute of Cognitive Neuroscience in Leipzig. (Technical details as regards the MRI machine are described in Brass and von Cramon, 2002.)

Design

The independent variables of the experiment were trial type (task shift vs repeat), cue-target interval (100 vs 2000 ms), and stimulus valence (bivalent vs univalent). Trial type and cue-target interval were varied randomly within a block, whereas stimulus valence was varied between blocks of trials. Group was a between-subject variable (standard behavioral, simulated MRI, and real MRI). Dependent variables were RT and error rates. All statistical analyses were run with an alpha level of 0.05.

Results and discussion

We do not report the brain-imaging data in this paper, but focus on the behavioral data. The two practice phases preceding each experimental block were not analyzed. For the analysis of the experimental blocks, we first discarded extremely slow responses (RT > 1500 ms; 0.5% of all responses). For RT analysis, we also discarded trials with incorrect responses (for error rates, see below). We then determined the mean RT for each participant as a function of trial type, cuing interval, and stimulus valence. Finally, we averaged the individual cell means for each group of participants (see Table 1).

We submitted the RT data to a four-way analysis of

variance (ANOVA) with the independent variables trial type, cuing interval, stimulus valence, and group. This yielded significant main effects of trial type, $F(1, 49) = 140.741$, $P < 0.001$; cuing interval, $F(1, 49) = 35.342$, $P < 0.001$; and stimulus valence, $F(1, 49) = 121.843$, $P < 0.001$. Stimulus valence interacted significantly with trial type, $F(1, 49) = 39.793$, $P < 0.001$, and with cuing interval, $F(1, 49) = 42.836$, $P < 0.001$. With univalent stimuli, there were shift costs of 18 ms, but shift costs were 48 ms with bivalent stimuli. The cuing interval effect was only 6 ms with univalent stimuli, but it was 76 ms with bivalent stimuli. The cuing interval also affected shift costs, $F(1, 49) = 43.598$, $P < 0.001$, indicating that shift costs were 52 ms with the short cuing interval but only 13 ms with the long cuing interval. Finally, there was a significant three-way interaction of trial type, stimulus valence, and cuing interval, $F(1, 49) = 23.635$, $P < 0.001$, indicating that long preparation time reduced shift costs in the univalent condition by 16 ms (i.e., from 25 to 9 ms), but that the shift-cost reduction was much stronger in the bivalent condition (62 ms; i.e., from 79 to 17 ms).

Together, these effects conform very well to the predictions that we derived from the task-switching model proposed by Meiran (2000). Providing long preparation time serves to reduce shift costs, suggesting advance reconfiguration of task set. The stimulus-related aspect of this reconfiguration process was expected to be weaker with univalent stimuli, because such stimuli do not require reconfiguration of stimulus set. In line with this prediction, we found that both shift costs and preparation effects were very small in the univalent condition as compared with the bivalent condition. In sum, across all three groups of participants, we obtained the pattern of results that we predicted based on our theoretical assumptions.

The crucial question of the present study was whether this complex and nontrivial data pattern would be the same across all groups or whether it would differ as a function of testing environment. In order to answer this question, the effects of the between-subject variable group are relevant. In fact, we got a significant main effect of group, $F(1, 49) = 8.707$, $P < 0.001$. Mean RT level was 452 ms (SE = 22 ms) in the standard behavioral group, 513 ms (SE = 22 ms) in the simulated MRI group, and 586 ms (SE = 23 ms) in the real MRI group. Pairwise comparisons indicated significant differences between each pair of groups (all P s < 0.05), statistically confirming that participants in the behavioral group were fastest, those in the real MRI group slowest, and those in the simulated MRI group in between. Importantly, apart from this main effect of group, there were no significant interactions of the group variable with any one, or any combination, of the other three variables (all P s > 0.23), indicating that the effects of preparation, trial type, and valence as well as their interactions were statistically identical in the three groups.

The pattern of statistical effects in the error rates closely mirrors that found for the RT, so we do not report these data

in detail. We observed the same statistical effects in the ANOVA, except that the three-way interaction of trial type, cuing interval, and stimulus valence failed to reach significance, $F(1, 49) = 3.191$, $P = 0.080$. Also the main effect of group, $F < 1$, and all interactions with group were insignificant, all P s < .14. Hence, the general RT slowing in the simulated MRI group and even more so in the real MRI group as compared to that in the standard behavioral group is not counteracted by more accurate responses, thus ruling out a simple speed-accuracy trade-off account for the present RT data pattern.

In order to evaluate the validity of the significant main effect of group in the RT data, we note that the participants in the real MRI group came from a different pool of participants than those in the other two groups. Also, they were tested in Leipzig, whereas the other two groups were tested in Munich. Therefore, one might argue that preexperimental group differences might partially explain the main effect of group. However, the RT data show that also the two groups tested in Munich differ significantly in RT level, which is to be explained by the experimental variation. Therefore, we think it is very likely that also the slowing in the real MRI group is due to the experimental conditions. In fact, given that potential distracting effects of MRI conditions should be stronger in the real MRI than in a simulation, then this pattern of results is exactly what one would have expected.

To summarize, we found a significant influence of the experimental setting on general RT level, but this influence was nonspecific and did not interact with any other variable. Based on this finding, we conclude that our data support the idea that cognitive control processes in task switching are functionally equivalent inside and outside the scanner. Note that, strictly speaking, this empirical conclusion is limited to the domain of task switching. However, it is clearly important to know whether this functional equivalence applies to other research domains as well. This raises the issue of which processes are actually affected inside the scanner that produce the general slowing that we observed in our study.

To explain this general slowing, we believe that unspecific distracting factors inside the scanner, such as the supine position and the noise, delay some aspect of performance. In order to specify this aspect, we can first conclude on the basis of our data that cognitive control processes (i.e., reconfiguration) are not affected because shift costs, as empirical marker for these processes, were virtually the same across all testing conditions. Hence, the unspecific distraction appears to prolong perceptual or motor components. We believe that it is highly unlikely that early visual processes (e.g., feature extraction) are affected, but more perceptual-cognitive aspects (e.g., stimulus identification) might be a possible candidate because, in the present experiment, the “up-down” task might be recoded as “in front-behind” when being in a supine position. However, this would be true only for the vertical dimension, whereas processing of the horizontal dimension (the left-right task) should be unchanged. We tested this possibility by running

an ANOVA that additionally included the independent variable task (up–down vs left–right). We found that the left–right task was performed on average 26 ms faster than the up–down task ($P < 0.001$), but that no interaction involving the task variable and the between-subject variable (i.e., testing condition) was significant. Therefore, we believe that this perceptual–cognitive aspect is unlikely to account for the general slowing.

Other possible cognitive aspects of information processing to account for the slowing inside the scanner might be categorization and response selection. Our valence manipulation appears to tap these processing aspects, because in the bivalent condition, stimulus selection, stimulus–response transformation, and response selection have to proceed in the face of a distractor (such as in the Stroop color-word task, see MacLeod, 1991, for review). Similar processes may underlie stimulus–response compatibility effects (for review, see Kornblum et al., 1990; Lien and Proctor, 2002). However, our data clearly indicate that this aspect is the same across testing conditions (i.e., no interaction including the valence variable and the experimental group variable was significant), so it cannot account for the slowing.

Hence, based on this analysis we can tentatively rule out the notion that perceptual and cognitive processes (e.g., reconfiguration, stimulus identification, response selection) are subject to the general response delay inside the scanner. This renders “late” processes of motor initiation and execution the most likely candidate to account for this slowing. One speculative account of this “motor” slowing might be that participants do not see their hands inside the scanner, which might increase response monitoring. However, alternatively it might be that still unknown physiological factors might play a role. Given the practical relevance of knowing the exact causes of the observed response slowing inside the scanner, we believe that it would be desirable to explore these causes also in experimental paradigms different from the one used here. Hence, we would like to encourage other cognitive neuroscience researchers to conduct similar studies with their experimental paradigm.

With respect to our conclusion of functional equivalence of cognitive processes inside and outside the scanner, note that we explicitly tested only the effects of rather unspecific factors associated with the testing environment in brain-imaging studies. There might be other factors affecting the functional equivalence of cognitive processes across studies that are related more specifically to aspects of the experimental design. For instance, in order to run an event-related fMRI study, one normally introduces jitter in the onset of events, which is typically avoided in behavioral studies. Perhaps even more important, event-related fMRI studies typically use intertrial intervals longer than those in behavioral studies. This might be an important difference across brain-imaging and behavioral studies in the context of task switching, because it has been argued that “decay” of the preceding task set during the intertrial interval might play

some role (Koch, 2001; Meiran et al., 2000). However, in the context of the present research, we can exclude such factors (e.g., the length of the intertrial interval) as possible sources of the general RT slowing we observed inside the scanner because we kept these factors strictly constant across our three experimental groups. In fact, the novel aspect of our study is that we investigated the effect of testing condition under otherwise identical experimental conditions, whereas neuroimaging studies typically establish in their experimental paradigm the behavioral effect they are interested in but do not compare it to that found in standard testing conditions. In our study, we matched these conditions as closely as possible. Our results suggest that the effect of testing condition is most likely functionally located at “late” motor processes and that cognitive processes are unaffected by the testing environment (i.e., inside vs outside the scanner).

Still, despite our conclusion that, most likely, cognitive processes inside and outside the scanner are functionally equivalent, our finding of greatly elevated baseline RT inside the scanner has potentially important methodological implications. Consider, for instance, experiments requiring a response deadline. Researchers using such techniques should keep in mind that strict response deadlines that are established in behavioral pilot experiments are unlikely to be appropriate during the scanning session because baseline RT will be increased. As another example, consider experiments on stimulus–response compatibility using Stroop-like, Simon-like, or flanker interference tasks, in which nominally irrelevant stimulus features interfere with performance (for review, see, e.g., Kornblum et al., 1990; Lu and Proctor, 1995). Here, the relevant cognitive process underlying the usually observed compatibility effect is assumed to depend on a temporally decaying code (Hommel, 1994; Kornblum et al., 1999). This code decay might progress to a larger degree due to the longer RT in the scanner, so that the general slowing inside the scanner relative to normal behavioral testing conditions could affect the size of the effect. However, our results suggest that the general slowing inside the scanner presumably does not affect cognitive processes such as response selection, which is commonly held responsible for compatibility effects (cf. Lien and Proctor, 2002, for review), but “late” motor processes subsequent to response selection, so that the cognitive process under study is most likely unaffected. Still, we believe that this issue represents an important topic for further investigation. We think that drawing the attention of cognitive neuroscience researchers using brain-imaging techniques to this potentially relevant methodological issue is an important contribution of the present study.

Acknowledgments

This project was supported by Grant G-635-88.4/1999 of the German-Israeli Foundation for Scientific Research and

Development (G.I.F.). The authors would like to thank the reviewers for their helpful comments on an earlier version of this article.

References

- Allport, A., Styles, E.A., Hsieh, S., 1994. Shifting intentional set: exploring the dynamic control of tasks, in: Umiltà, C., Moscovitch, M. (Eds.), *Attention and Performance*, Vol. XV, Conscious and Nonconscious Information Processing. MIT Press, Cambridge, MA, pp. 421–452.
- Brass, M., von Cramon, D.Y., 2002. The role of the prefrontal cortex in task preparation. *Cereb. Cortex* 12, 908–914.
- D’Esposito, M., Zarahn, E., Aguirre, G.K., 1999. Event-related functional MRI: implications for cognitive psychology. *Psychol. Bull.* 125, 155–164.
- Dove, A., Pollmann, S., Schubert, T., Wiggins, C.J., von Cramon, D.Y., 2000. Prefrontal cortex activation in task switching: an event-related fMRI study. *Cognit. Brain Res.* 9, 103–109.
- Hommel, B., 1994. Spontaneous decay of response-code activation. *Psychol. Res.* 56, 261–268.
- Kimberg, D.Y., Aguirre, G.K., D’Esposito, M., 2000. Modulation of task-related neural activity in task-switching: an fMRI study. *Cognit. Brain Res.* 10, 189–196.
- Koch, I., 2001. Automatic and intentional activation of task sets. *J. Exp. Psychol. Learn. Mem. Cognit.* 27, 1474–1486.
- Kornblum, S., Hasbroucq, T., Osman, A., 1990. Dimensional overlap: cognitive basis for stimulus–response compatibility: a model and taxonomy. *Psychol. Rev.* 97, 253–270.
- Kornblum, S., Stevens, G.T., Whipple, A., Requin, J., 1999. The effects of irrelevant stimuli. 1. The time course of stimulus–stimulus and stimulus–response consistency effects with Stroop-like stimuli, Simon-like tasks, and their factorial combinations. *J. Exp. Psychol. Hum. Perception Performance* 25, 688–714.
- Lien, M.-C., Proctor, R.W., 2002. Stimulus–response compatibility and psychological refractory period effects: implications for response selection. *Psychonomic Bull. Rev.* 9, 212–238.
- Lu, C.-H., Proctor, R.W., 1995. The influence of irrelevant location information on performance: a review of the Simon and spatial Stroop effects. *Psychonomic Bull. Rev.* 2, 174–207.
- MacLeod, C.M., 1991. Half a century of research on the Stroop effect: an integrative review. *Psychol. Bull.* 109, 163–203.
- Meiran, N., 1996. Reconfiguration of processing mode prior to task performance. *J. Exp. Psychol. Learn. Mem. Cognit.* 22, 1423–1442.
- Meiran, N., 2000. The reconfiguration of the stimulus task-set and the response task-set during task switching, in: Monsell, S., Driver, J. (Eds.), *Attention & Performance*, Vol. XVIII, Control of Cognitive Processes. MIT Press, Cambridge, MA, pp. 377–400.
- Meiran, N., Chorev, Z., Sapir, A., 2000. Component processes in task switching. *Cognit. Psychol.* 41, 211–253.
- Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202.
- Rogers, R.D., Monsell, S., 1995. Costs of a predictable switch between simple cognitive tasks. *J. Exp. Psychol. General* 124, 207–231.
- Sohn, M.-H., Ursu, S., Anderson, J.R., Stenger, V.A., Carter, C.S., 2000. The role of prefrontal cortex and posterior parietal cortex in task switching. *Proc. Natl. Acad. Sci. USA* 97, 13448–13453.