Real-time fMRI neurofeedback: progress and challenges

Sulzer, J; Haller, S; Scharnowski, F; Weiskopf, N; Birbaumer, N; Blefari, M L; Bruehl, A B; Cohen, L G; DeCharms, R C; Gassert, R; Goebel, R; Herwig, U; LaConte, S; Linden, D; Luft, A; Seifritz, E; Sitaram, R

Abstract: In February of 2012, the first international conference on real time functional magnetic resonance imaging (rtfMRI) neurofeedback was held at the Swiss Federal Institute of Technology Zurich (ETHZ), Switzerland. This review summarizes progress in the field, introduces current debates, elucidates open questions, and offers viewpoints derived from the conference. The review offers perspectives on study design, scientific and clinical applications, rtfMRI learning mechanisms and future outlook.

DOI: https://doi.org/10.1016/j.neuroimage.2013.03.033

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: https://doi.org/10.5167/uzh-85493

Accepted Version

Originally published at:
Sulzer, J; Haller, S; Scharnowski, F; Weiskopf, N; Birbaumer, N; Blefari, M L; Bruehl, A B; Cohen, L G; DeCharms, R C; Gassert, R; Goebel, R; Herwig, U; LaConte, S; Linden, D; Luft, A; Seifritz, E; Sitaram, R (2013). Real-time fMRI neurofeedback: progress and challenges. NeuroImage, 76:386-399.

DOI: https://doi.org/10.1016/j.neuroimage.2013.03.033
Real-time fMRI Neurofeedback: Progress and Challenges

J Sulzer\textsuperscript{a*}, S. Haller\textsuperscript{b*}, F. Scharnowski\textsuperscript{c,d*}, N. Weiskopf\textsuperscript{e*}, N. Birbaumer\textsuperscript{f,g}, M.L. Blefari\textsuperscript{a}, A. Bruehl\textsuperscript{h}, L. G. Cohen\textsuperscript{i}, R C deCharms\textsuperscript{j}, R. Gassert\textsuperscript{k}, U. Herwig\textsuperscript{h}, S. LaConte\textsuperscript{l}, D. Linden\textsuperscript{m}, A. Luft\textsuperscript{n}, E. Seifritz\textsuperscript{h}, R. Sitaram\textsuperscript{f,o,p*}

*corresponding authors

\textsuperscript{a}Department of Health Sciences and Technology, Swiss Federal Institute of Technology, Zurich, Switzerland (ETH), CH-8092

\textsuperscript{b}University of Geneva, Geneva University Hospital, Switzerland, CH-1211

\textsuperscript{c}Department of Radiology and Medical Informatics - CIBM, University of Geneva, Switzerland

\textsuperscript{d}Institute of Bioengineering, Swiss Institute of Technology Lausanne (EPFL), Switzerland, CH-1015

\textsuperscript{e}Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, University College London

\textsuperscript{f}The Institute of Medical Psychology and Behavioral Neurobiology, University of Tuebingen, Germany, 72074

\textsuperscript{g}Ospedale San Camillo, IRCCS, Venice, Italy, 30126

\textsuperscript{h}Psychiatric University Clinic, Zurich, Switzerland, CH-8032

\textsuperscript{i}National Institutes of Health, Bethesda, USA 20892

\textsuperscript{j}Omneuron, Inc., Menlo Park, USA 94025

\textsuperscript{k}Department of Neurocognition, University of Maastricht, The Netherlands, 6200

\textsuperscript{l}Virginia Tech Carilion Research Institute, USA 24016

\textsuperscript{m}Cardiff University, UK CF10 3AT

\textsuperscript{n}Department of Neurology, University Hospital Zurich and University of Zurich, CH-8008
Corresponding Authors

James Sulzer, Leonhardstrasse 27 B9.2, Zurich Switzerland, CH-8092, +41 44 632 6763,
jamessulzer@gmail.com

Sven Haller, Rue Gabrielle Perret-Gentil 4, Geneva, Switzerland CH-1211, +41 79 553 2528,
sven.haller@hcuge.ch

Frank Scharnowski, IBI-STI,GRVDV, BM4.129, Station 17, Lausanne, Switzerland CH-1015,
+41 78 767 6749, scharnowski.frank@gmail.com

Nikolaus Weiskopf, 12 Queen Square, London WC1N 3BG, UK, +44 20 3448 4381,
n.weiskopf@ucl.ac.uk

Ranganatha Sitaram, Department of Biomedical Engineering, University of Florida, Gainesville,
USA 32611.
Abstract

In February of 2012, the first international conference on real time functional magnetic resonance imaging (rtfMRI) neurofeedback was held at the Swiss Federal Institute of Technology Zurich (ETHZ), Switzerland. The results of this conference inspired the idea to disseminate current state-of-the-art, but also to delineate the open areas of research. This review summarizes the progress in the field, introduces current debates, elucidates open questions, and offers viewpoints. We offer this analysis from the perspectives of study design, scientific and clinical applications, its learning mechanisms and future outlook.
0 Introduction

On February 16th and 17th, 2012, approximately 150 international researchers joined the first conference on an emerging discipline known as real-time functional magnetic resonance imaging (rtfMRI) neurofeedback at the Swiss Federal Institute of Technology in Zurich (ETHZ), Switzerland (www.relab.ethz.ch/rtfMRI2012). The purpose of this meeting was to provide a forum to share progress and discuss the challenges for future research and clinical applications. The meeting also inspired the creation of the following work, which reviews current progress and introduces open questions and controversies.

Functional MRI measures the blood oxygenation level dependent (BOLD) signal in the brain (Ogawa et al. 1990a; Ogawa et al. 1990b), a quantity that arises from several biophysical and physiological sources (Kim and Ogawa 2012) and represents a vascular coupling to neural activity (Logothetis 2008; Logothetis et al. 2001). Despite large size, cost and rather low temporal resolution, fMRI has specific advantages over other non-invasive neuroimaging methods such as electroencephalographic recordings (EEG) including whole brain coverage and finer spatial resolution on the order of one millimeter. We define rtfMRI, first published by Cox and colleagues (Cox et al. 1995), as any process that uses functional information from a MRI scanner while the scan is being conducted. To do this, fMRI volumes can be processed through direct software access, or on a remote computer via network transmission, or through a shared network drive. Although whole brain fMRI sampling periods can now be performed at around half a second (Feinberg et al. 2010), typical protocols still use rates of approximately every two seconds. Since these rates are relatively slow and because tight integration with MRI hardware is
vendor specific, most rtfMRI setups access the image data via a shared drive. Cox et al. described that real-time brain mapping could be used for quality assurance, faster protocol development and "interactive experimental paradigms". At present, rtfMRI has additionally been applied to intraoperative surgical guidance (Hirsch et al. 2000), brain-computer interfaces (BCIs) (Yoo et al. 2004), and neurofeedback.

While EEG neurofeedback has a long history (Elbert et al. 1980; Rockstroh et al. 1984; Rockstroh et al. 1993), there has been a recent rise in attention to rtfMRI neurofeedback, providing a timely background for the conference. Figure 1 shows that there were more journal papers published on the topic in 2011 (38) than the previous four years combined (36). The figure illustrates that recently neurofeedback and methods development currently comprise the majority of the field, and, as a result, this paper focuses on neurofeedback approaches (Berman et al. 2011a; Bray et al. 2007; Caria et al. 2010; Caria et al. 2007; Chiew et al. 2012; deCharms et al. 2004; deCharms et al. 2005; Frank et al. 2012; Haller et al. 2010; Hamilton et al. 2011; Hawkinson et al. 2011; Hawkinson et al. 2012; Hinds et al. 2011; Johnston et al. 2010; Johnston et al. 2011; Lee et al. 2012; Lee et al. 2011; Li et al. 2012; Linden et al. 2012; McCaig et al. 2011; Posse et al. 2003; Rota et al. 2009; Shibata et al. 2011; Subramanian et al. 2011; Veit et al. 2012; Weiskopf 2011; Weiskopf et al. 2004a; Weiskopf et al. 2003; Yoo and Jolesz 2002; Yoo et al. 2008; Zotev et al. 2011). Figure 1 also clearly shows that review papers regarding this technology are rather plentiful (e.g. (Caria et al. 2012; Chapin et al. 2012; deCharms 2008; LaConte 2011; Linden 2012b; Sitaram et al. 2010; Weiskopf et al. 2004b)). Therefore the purpose of this paper is to focus more on the open questions indentified during the conference and the challenges that lie ahead. The paper is divided into five subsections that
examine rtfMRI neurofeedback from different perspectives: 1) study design, 2) scientific applications, 3) clinical applications, 4) learning mechanisms and 5) the future of rtfMRI neurofeedback.

1 Considerations in Study Design

The design of a study depends on its objectives. The experimental objectives of neurofeedback studies may range from demonstrating neurofeedback induced learning of self-regulation to specific behavioral effects (e.g. (Rota et al. 2009; Shibata et al. 2011)) or clinical improvement in patients (e.g. (deCharms et al. 2005; Ruiz et al. 2011; Subramanian et al. 2011)). However, the majority of neurofeedback studies employ a similar experimental framework and schedule, primarily consisting of:

1. Definition of the physiological target and response: a region is anatomically specified or a functional localizer is applied to define the brain region, network and/or physiological response to be trained.

2. Neurofeedback of the physiological target response: the participant is presented with feedback of the physiological target to be trained (see Figure 2). Feedback training may span several minutes, hours, or repeated sessions over days.

3. Transfer after successful training: when the participants have achieved successful regulation, one needs to test whether they are able to maintain the skill in the absence of feedback and/or in a different setting or task.

4. Experimental control: studies employed different control groups or within subject control conditions to control for confounds in learning, behavioral and placebo effects.
5. Testing of behavioral effects: after participants learned effective regulation, one can test if this results in specific behavioral effects before and after learning.

1.1 Definition of the physiological target and response

The definition of the neurofeedback target typically depends on the behavioral effect that should be achieved. For example, experiments that aimed at modulating reaction times, manipulated the activity in motor areas such as the supplementary motor area (SMA) or primary motor cortex (M1) (Bray et al. 2007; Weiskopf et al. 2004a). Another experiment, which aimed at changing pain perception, regulated activity in the rostral anterior cingulate cortex (rACC) (deCharms et al. 2005). Shibata et al. aimed at specific voxels in primary and secondary visual cortices to evoke a change in visual perception (Shibata et al. 2011). Using previous knowledge of neural mechanisms underlying the desired behavioral change is the key to selecting the desired ROI.

The physiological target may be the average BOLD response in a chosen ROI, but it may also be more complex such as the differential activity in two ROIs (Chiew et al. 2012; Weiskopf et al. 2004b) or activity in a multi-region network (LaConte et al. 2007)). While a mean BOLD response of a ROI is the most straightforward and easily interpreted signal, the differential BOLD response from two different regions may offer more control over unspecific physiological effects (Fox and Rudell 1968). For example, breathing artifacts should cancel out because they have similar effects on both target regions. On the other hand, while some noise may be correlated, the uncorrelated Gaussian noise of the two signals will be additive, thus reducing the signal-to-noise ratio. Multivariate pattern analyses (MVPA) of BOLD responses allow the
experimenter to identify complex and interacting activity patterns, probably best reflecting network activity (LaConte et al. 2007). Ongoing studies explore the possibility for feedback of connectivity between brain areas (e.g. presentations by Ruiz, and Zilverstand), similar to functional or effective connectivity measures used off-line (Friston et al. 2003; Roebroeck et al. 2005). Specific ROIs can often be anatomically defined based on brain atlases or macroscopic anatomical landmarks, such as the insular cortex (Caria et al. 2007) but also functionally defined, such as the parahippocampal place area (Weiskopf et al. 2004a). A combination of overlaying functional activity on anatomical images may help further improve demarcation (e.g. hand knob of primary motor cortex presented by Blefari). Brain networks are usually difficult to define anatomically due to high variability. However, anatomical localizers may be more appropriate in certain cases where the anatomical region is well defined and a reliable functional localizer is difficult (e.g. substantia nigra presented by Sulzer). Some unpublished evidence comparing functional to anatomical localizers for a given ROI was offered at the conference, showing that functional localizers offer a better contrast-to-noise signal, but that head movement of one millimeter or greater removes this advantage over anatomical selection (presentation by deCharms).

1.2 Neurofeedback of physiological target response

Participants are trained by providing feedback of the previously defined physiological target response. In previous studies feedback was mostly presented visually as a thermometer reading or scrolling curve. However, feedback was also implemented via virtual reality, such as reaching for a coffee mug (Sitaram et al. 2005) or computer games (Goebel et al. 2004a). During the meeting, the impact of neurofeedback interfaces and how to potentially evaluate and optimize
their design was noted as a current research gap in the field. In addition, the conference
participants discussed a frequent lack of methodological detail in articles, making it difficult for
other groups to replicate studies using the same processing parameters and rules for feedback
display updates (i.e. methods for calculating percent signal change and descriptions of how
measured changes are related to number of units and visual field angles incremented or
decremented from a thermometer).

In the great majority of studies feedback was continuously presented with minimal delay,
approximately every 2 s depending on the volume acquisition rate, which may be a result of
adaptation from EEG neurofeedback studies (e.g. (Kotchoubey et al. 2001)). Alternatively, in
some studies feedback was presented after longer blocks of up to one minute (Bray et al. 2007;
Posse et al. 2003; Shibata et al. 2011; Yoo and Jolesz 2002) and one study reported improved
learning in intermittent feedback compared to continuous feedback (Johnson et al. 2012).
Section 4: Learning Mechanisms discusses the reasoning behind this in more detail.

Briefing and debriefing of the participants can be an important part of neurofeedback training
unique to cognitively-intact human subjects (Birbaumer et al. 2008). Typically the feedback
signal and its delay with respect to neuronal activity are explained to the participants (deCharms
et al. 2005; Yoo and Jolesz 2002). Some studies also did short pre-trainings with computer-aided
programs to acquaint participants with the delay of the feedback due to the hemodynamic
response and the computing time for the feedback signal (presented by Hollman). Volunteers are
often instructed to minimize head motion and irregular breathing, in order to minimize a
systematic influence of physiological artifacts on the feedback signal (Zhang et al. 2011a). It
may also be explained that the signal is relatively noisy, in order to manage the subject's
expectations of perfect control. With few notable exceptions, e.g. (Kim et al. 2011; Shibata et al.
202011) the majority of studies explicitly suggested control strategies to the participants, such as
imagery or attention strategies known to be related to the targeted physiological process and
area. Most experimenters encourage the volunteers to develop their individual strategies from
these initial ones. Several studies also included (interim) debriefing of participants including
subjective reports of success and control strategies (Shibata et al. 2011; Sitaram et al. 2011).

A topic of debate at the conference was the primacy of implicit or explicit strategies. Explicit
strategies entail informing the subject of a specific means for self-regulation, whereas implicit
strategies provide no such instruction and allow the subject greater room to explore. Implicit
strategies involve no suggested instructions regarding any particular strategy to use for self-
regulation. On one hand, implicit strategies (e.g. (Shibata et al. 2011)) may be better since
compliance to a suggested cognitive strategy cannot be quantitatively confirmed, and it may be
difficult for some people to understand or report. Additionally, such specification limits the set
of possible optimal imagery strategies. The conference participants noted that there is no
published report that directly compares implicit and explicit strategies in terms of their
effectiveness in learning to self-regulate brain regions. However, one unpublished study
supporting explicit strategies examined neurofeedback of the language area in 16 participants
first using implicit strategies and finding no learning (presented by deCharms). When subjects
were then given the explicit strategies, they were able to learn the task. However, it may be that
implicit learning takes longer or possibly that this effect may be specific for language learning
and not simply transferable to other tasks. There are also many other considerations to account
for in this unresolved debate, such as some regions may have no associated explicit strategy,
costly and limited scanner time, and the specific hypothesis to be tested. Indeed, until the
mechanisms behind such learning are better understood (see Section 4: Learning Mechanisms), it
may be difficult to reach a conclusion.

Multivariate pattern analysis (MVPA)-based rtfMRI, on the other hand, tends to be explicitly
task-based; instead of anatomically or functionally localizing ROIs, classifiers use task
conditions during a training step. Thus, instead of implicit or explicit strategies, for modulating
ROIs, the focus is on what brain networks best discriminate the different task conditions for a
particular subject and their individual cognitive strategy (LaConte 2011).

Most studies employ a block design for the regulation task. In this type of design, volunteers are
required to regulate the BOLD signal for usually 15-30s followed by a rest block of similar
duration. Unlike event-related designs, block designs are less sensitive to undesired delays due to
the required task switching and slow BOLD response. A single run consists of 3-6 blocks, lasts
ca. 5-15 minutes and is repeated 2-5 times within an experimental session. The number of
sessions varies significantly between studies from a single session to up to 10 sessions (Shibata
et al. 2011), but the majority consisted of a single session (e.g. (Caria et al. 2007; deCharms et al.
2005)). It is not clear why in some experiments learning curves were much steeper than in others,
resulting in such a widespread difference in duration. Systematic studies on the optimal duration
of runs, repetition of sessions and gaps between sessions are lacking, although anecdotal
evidence from the field of perceptual learning suggests that shorter runs are more effective
(Molloy et al. 2012). Offline mental training between sessions could be advantageous towards
accelerating learning (Subramanian et al. 2011), but make it difficult to separate its effect from neurofeedback training. When explicit strategies are suggested, anecdotal evidence was offered that offline coaching by the experimenters could also have a positive effect on performance (presentation by deCharms). The maximal number of runs seems to be limited by the attention span and exhaustion of volunteers.

Although the majority of studies use the same fixed training duration for all subjects, it has been recognized that adaptive designs may be more appropriate, since the individual learning curves can vary significantly. Thus, at least one study introduced individual criteria for finalizing the training (Scharnowski et al. 2010), e.g., based on achieved success of regulation in transfer trials (see Section 3: Clinical Applications). This may help to make group effects more homogeneous in following behavioral tests.

Typically, neurofeedback is conducted without any explicit stimulation, although this need not be the case. For instance, Veit and colleagues (Veit et al. 2012) trained participants to volitionally up- and down-regulate the anterior insula in the presence of threat-related stimuli. Another study employed down-regulation of amygdala during emotional stimuli (presented by Bruehl). Yet another study trained individuals to up- and down-regulate, in separate sessions, brain regions involved in the visual perception of emotion, when subjects were concurrently stimulated by a backward priming paradigm ((Kim et al. 2011), presentation by Sitaram). It is not clear how a stimulus-based self-regulation may be more preferable in terms of maintenance of attention, block/run length and other study design parameters.
1.3 Transfer after successful training

Since most studies aim at investigating behavioral effects, it is crucial that participants maintain self-regulation in the absence of feedback and outside the scanner. In particular, in clinical applications an important goal will be to maintain skills practiced and acquired during rtfMRI sessions and be able to apply them to real-life situations. Most studies included transfer runs that followed the same experimental design as training runs but lacked the feedback signal (e.g. Caria et al. 2007; deCharms et al. 2004; deCharms et al. 2005; Ruiz et al. 2011)). Usually transfer runs are conducted at the end of an experimental session or after a number of sessions. As transfer should demonstrate the degree to which the learned regulatory ability can be translated to the world outside the scanner, some studies use a similar, but different paradigm compared to the training task (Caria et al. 2007; Sulzer et al.). Debriefing after transfer sessions have used subjective reports of regulation success to assess placebo effects and awareness – similar to EEG feedback studies (Kotchoubey et al. 2001). More clinically-oriented studies will likely desire long-term monitoring of behavioral consequences, adding a follow-up behavioral evaluation long after rtfMRI training (See Section 3: Clinical Applications).

1.4 Experimental control conditions

The experimental controls employed in rtfMRI neurofeedback can serve various purposes. In most cases they were used to determine whether the feedback itself is necessary for learning the self-regulation compared to simple instructions alone. Control groups received sham feedback that was derived from other participants’ data or artificially created (Caria et al. 2010; deCharms...
et al. 2004; deCharms et al. 2005; Rota et al. 2009). In other studies control groups received contingent feedback (i.e. directly related to the feedback signal), but from areas other than the experimental target region (deCharms et al. 2005; Scharnowski et al. 2010), which can control for psychoeducative (i.e. benefit from learning) effects. In another study, feedback was inverted to encourage down-regulation of the selected ROI, which can more strongly show a differential effect of neurofeedback (presented by Sulzer). In the sham feedback paradigm the success rates can be well-matched between the experimental and control group, but it may not present a realistic feedback with respect to noise and contingency. However, a subject may consciously or unconsciously interpret the less representative sham or control region feedback, thereby discouraging performance and creating a placebo expectancy effect (Stroebel and Glueck 1973).

In some studies, regulation without the feedback has been used as a control condition (deCharms et al. 2005).

The control groups mentioned above are all examples of negative controls, i.e. conditions that one would expect to show worse performance than the experimental group. Alternatively, in some cases, employing a positive control group can provide a more ecological comparison. For instance, if one were to evaluate the usefulness of neurofeedback in selectively activating a target ROI, it should be compared to the best-known method of exciting that region. This was the strategy used by Berman et al. who examined self-regulation of primary motor cortex, finding that self-regulation using finger tapping exhibited, as expected, far superior performance to that of mental imagery strategies (Berman et al. 2011b). A positive control may even be in the same run, for example, during stimulation, down-regulation of amygdala compared to the activity level during passive viewing (presentation by Bruehl). Inclusion of positive controls is a necessary
An exciting and emerging focus for many groups using rtfMRI neurofeedback has moved from learning regulation to testing specific behavioral effects. Thus, the experimental controls now aim at controlling for confounds in behavioral tests. Sham feedback and contingent feedback from an alternative area were used to test for specificity (deCharms et al. 2005). However, particularly in clinical studies new possibilities and issues arise. In these studies control groups who receive a completely different type of treatment were introduced to control for placebo effects and estimate the relative efficiency, since it may be less important to estimate the precise effect of the feedback. For example, in a study on chronic pain, rtfMRI neurofeedback was compared to skin conductance response feedback (deCharms et al. 2005). In Parkinson’s disease, neurofeedback was compared to motor imagery (Subramanian et al. 2011). Within subject controls are also possible by training two mutually exclusive physiological responses. An example is the bidirectional regulation of the BOLD response. For example, in such a bidirectional control design, significantly different memory encoding effects were shown for the up- vs. down-regulation condition (Weiskopf et al. 2004b). Such an internal control reins in on unspecific attention and regulation effects and does not require matching of different groups. Using a different strategy, Shibata and colleagues trained each subject on one of three different grating patterns and found differential improvements based on a functionally localized ROI (Shibata et al. 2011). Placebo effects can also be controlled for by subjective reports as shown in...
EEG-feedback (Kotchoubey et al. 2001; Schwartz and Andrasik 2003) but this has not yet been implemented in rtfMRI feedback.

Signal artifacts can contaminate BOLD measurements. A recent example of this was demonstrated by (Zhang et al. 2011b), who showed that eye movements could inflate rtfMRI training effects in the slices limited to where the eyes are recorded. Physiological noise from sources such as heart rate and respiration (Glover et al. 2000; Hu et al. 2005; Krüger et al. 2001) and head motion (Cox and Jesmanowicz 1999; Friston et al. 1996; Hajnal et al. 1994) are arguably the most relevant artifacts in fMRI. Many studies have employed online motion correction and some studies measured heart rate and breathing rates to control for systematic errors, and there are specific tools that are available that can be used in post-hoc analysis such as RETROICOR (Glover et al. 2000; Hutton et al. 2011; Kasper et al. 2009). Recent developments in signal processing in real-time fMRI can further improve the robustness against such unspecific effects and noise (Hinds et al. 2011; Koush et al. 2012).

In summary, there is currently no single "correct" experimental design in rtfMRI neurofeedback. While there are many basic elements that rtfMRI neurofeedback experiments have in common, experimental designs will vary depending on the specific hypothesis, ROI, behavior, and type of subject. As with most experiments, pilot testing is required to fine-tune various parameters, and to maximize learning and robustness. There still remain many fundamental open questions regarding optimization of designs, as noted in Box 1.
Neurofeedback as a scientific tool was pioneered by a number of researchers in the late 1960's (Fetz 1969; Fox and Rudell 1968; Olds 1965; Wyrwicka and Sterman 1968), using electrophysiological recordings in animals either noninvasively (EEG) or invasively. These research lines continue into the present time (Jackson et al. 2006; Moritz et al. 2008). In humans, a number of studies have demonstrated the feasibility of learning to control local brain activity using rtfMRI neurofeedback. Some of these studies have even shown that learned control of brain activity leads to behavioral effects that are specific to the functional role of the targeted brain area (Bray et al. 2007; Caria et al. 2007; deCharms et al. 2005; Haller et al. 2010; Rota et al. 2009; Shibata et al. 2011; Subramanian et al. 2011; Weiskopf et al. 2003). Whereas conventional neuroimaging shows simultaneous state or change of state of behavior and brain function, these neurofeedback results help reveal how changes in brain activity lead to changes in behavior or perception, i.e. brain activity is the independent variable. Yet as discussed at the conference, it is controversial whether neurofeedback can be used specifically as a tool for causal inference on specific neuronal mechanisms underlying behavior, which is the focus of this section.

In order to establish a causal link between brain activity and behavior, interventional manipulations have been proposed such as transcranial magnetic stimulation (TMS) (Tanaka et al. 2011), deep brain stimulation (Benabid et al. 1991), cortical cooling (Bauer and Fuster 1978), psychopharmacology (Angrist et al. 1980), or focal lesions in patients (Bhatia and Marsden 1994). They allow to study how manipulations of brain activity affect behavior and thus whether
a certain spatio-temporal activity of a brain region (and network) is a necessary component for a specific mental function (Censor et al. 2010). For example, when TMS is used to stimulate the visual cortex after the presentation of a visual stimulus, the stimulus will not be perceived (e.g. (Amassian et al. 1989)). This allows us to conclude that the spatio-temporal pattern of visual cortex activity was necessary for conscious perception of the visual stimulus. Although TMS may not be as focal as previously thought (Ruff et al. 2009), the stimulation method is independent of the quantity being measured, the location and magnitude of the stimulation is repeatable, and control conditions verify the effects of interest. Each tool for interventional manipulations have different strengths and drawbacks for causal inference, that can be evaluated with respect to their have different levels of independence, repeatability, controllability, and specificity.

In contrast to methods of exogenous (i.e. originating outside the body) stimulation or lesion, neurofeedback is based on endogenous (i.e. originating inside the body) manipulations of brain activity. In other words, neurofeedback trains participants to consciously or unconsciously modulate their own brain activity, equivalent to "self-stimulation". While learned voluntary manipulation of own brain activity may be an advantage for clinical applications, it limits control of dosage, is less precise than externally controlled stimulation, and makes it difficult to reproduce the exact same manipulation (i.e. controllability, specificity and repeatability). Yet despite these problems, causality with neurofeedback may not be out of reach. For instance, while not directly addressing causality, the previously mentioned clinical study by deCharms et al. examined how self-regulation of rACC correlated with pain perception (deCharms et al. 2005). Instead of trying to dose-match the self-regulation as with brain stimulation, the authors
correlated the ability to self-regulate rACC with reduced pain perception. The experiment also used three control groups experiencing no feedback, yoked sham feedback, and feedback from a different region, none of which showed the same effects as contingent rACC neurofeedback. These controls ensured that the effect did not arise from the explicit mental strategy given, observing rACC activity, or the ability to self-regulate any region, respectively. In terms of specificity, the authors admit that it is possible that rACC activity changes may be driven by top-down connections from a higher order region that causally affects both rACC activity and pain perception as independent quantities. Secondly, it may also be possible that the participants' abilities to self-regulate rACC may not be independent of the abilities to self-regulate pain.

Perhaps the most convincing case of demonstrating causality using rtfMRI neurofeedback thus far is from work on visual perceptual learning (VPL) by Shibata and colleagues (Shibata et al. 2011). In this study, they used a decoder to identify voxels in early visual cortex (V1/V2) corresponding to three different Gabor patch gratings differing by 60° orientation from each other. The feedback signal communicated the likelihood of these voxels representing one of the patches, unbeknownst to the participant. Following neurofeedback training, participants improved perceptual sensitivity to the target grating compared to the other two. These different gratings were an inventive way to establish control conditions separating the ability to self-regulate from behavioral effects. To account for specificity, the authors compare activity in other related regions offline to V1/V2 activity, showing that no other connected regions could account for this change.
In perhaps one way to overcome some inherent controllability and repeatability issues with endogenous neurofeedback, recent work has taken the approach of removing conscious human cognition from the control loop. For example, in an innovative study, Yoo and colleagues monitored the activation in a memory-related ROI in real-time and triggered a memory probe when participants entered good and bad brain states for learning novel scenes (Yoo et al. 2011). They found that when scenes were triggered by good ROI states, they were remembered significantly better than scenes that were triggered by bad ROI states. Hence, the activation patterns in the ROI were correlated with memory performance. However, from a causal perspective, the possibility of a higher order region or network being primarily responsible for this effect is not clear. Another example for a new real-time fMRI paradigm that does not need the active self-regulation of brain activity is the closed-loop paradigm. In such a paradigm, the sensory stimulation is modified depending on the current level of brain activity. For example, Gantner and colleagues changed the transparency of an image of a house depending on the level of activity in a house processing brain area (Gantner et al. 2010). The participants in this experiment were not aware that the visual stimulation is linked to their own brain activity. Such a closed loop paradigm can be used to investigate neuronal dynamics, where changes in brain activity cause changes in visual stimulation, which in turn causes changes in brain activity.

In the framework of a scientific tool, the rtfMRI neurofeedback has some known drawbacks, but also has a specific advantage over other interventions due to its whole brain coverage. As a result, neurofeedback studies have implemented functional connectivity-based (presentation by Zilverstand), multiple ROIs (Chiew et al. 2012), and machine learning classifiers (LaConte et al. 2007; Sitaram et al. 2011). While the application of TMS is limited to a direct stimulation of
cortical regions beneath the skull, rtfMRI has the ability to intervene virtually any combination
of brain areas. Table 2 contains a comparison of methods used for causal neuroscientific
investigation in humans.

These examples illustrate that rtfMRI-based paradigms significantly extend the possibilities of
conventional neuroimaging. In addition to expanding neuroscientific knowledge, if fMRI were to
become accepted as a clinical diagnostic tool, rtfMRI could play a role in helping identify causal
relations of the BOLD signal in neural function. Similar to other interventional techniques, the
new rtfMRI-based paradigms might allow us to address questions of causality rather than mere
correlations between brain activity and mental functions. However, important questions
regarding its use still remain, as summarized in Box 2.

3 Clinical Applications

Disorders of the brain, ranging from stroke to addiction to autism, represent one of the crucial
public health challenges for rtfMRI neurofeedback. The following section describes the steps to
be taken and risks to be considered if neurofeedback is to play a role in addressing this challenge.

Although a large variety of brain disorders could be imagined in principle as targets for
neurofeedback, robust and well-controlled studies on patients based on well-founded
pathophysiological models must lead the way. Until now, studies using rtfMRI neurofeedback
have shown that healthy subjects can self-regulate a number of different brain regions during
scanning. Less is known about the ability of patients with neurological and psychiatric disorders to learn self-regulation of focal brain activity through rtfMRI neurofeedback and the behavioral effects thereof. Previous literature in EEG neurofeedback has shown the ability to self-regulate brain activity in patients suffering from psychiatric and neurological disorders, including ADHD and epilepsy (Birbaumer et al. 2008). To date, six individual pilot studies in rtfMRI neurofeedback have reported training success with different patient groups (Table 1),

3.1 Which neural circuit to train?

One methodological aspect in the development of therapeutic rtfMRI neurofeedback is to differentiate between training aimed at improving deficient neural circuitries directly versus training “compensatory” circuits. The success of the former, which is akin to the approach generally taken in the development of deep-brain stimulation protocols, depends on sound knowledge of the disturbed circuits, especially when the model of the disorder is well-supported by multimodal evidence (animal studies, human studies, stimulation/lesion studies). The ROI or network targeted for rtfMRI neurofeedback should be accurately represented based on neuroscientific and clinical knowledge of the pathophysiology of the disorder at hand, which is a particular challenge in those psychiatric disorders where no clinically suitable imaging biomarkers have been identified (Linden 2012a).

If the specific biological mechanism is not well-known, an alternative would be to target potentially compensatory networks that have been well studied in the healthy population. One example of relevance to psychiatry are the putative networks for automatic and voluntary
emotion regulation (Phillips et al. 2008), which provide multiple targets for region- or network-based neurofeedback training (Esmail and Linden in press; LaConte et al. 2007; Sitaram et al. 2011). If we assume that patients have a clinical, psychological or cognitive deficit in a particular domain, for example in emotion regulation, it might make sense to engage them in a training process involving the relevant circuits without requiring demonstration of a primary deficit in these circuits. Although it may sound unsatisfactory to apply a treatment protocol without first demonstrating a biological deficit, this approach has been successfully implemented in nearly all psychiatric (both biological and psychological) therapeutics (Linden 2012a). However, one potential problem with this approach is that functional networks in patients may differ from those in the healthy population as a result of primary deficits or adaptations to the disease process. This problem can be addressed by identifying target areas through individual functional localizer scans. For example, a recent study on depression by some of the co-authors identified the brain areas responsive to positive emotional stimuli at the start of each neurofeedback session and used these functionally defined areas as ROIs for the self-regulation training (Linden et al. 2012).

3.2 Potential Risks

One of the first steps before clinical implementation is to evaluate the potential risks involved. When safety guidelines are properly followed, MRI and fMRI are regarded as relatively risk-free methods. In over 20 years of application, no severe adverse events have occurred or side effects have been detected as long as security guidelines are followed (Bourland et al. 1999; Schaefer et al. 2000; Schenck 2000; Shellock and Crues 2004). Though applied in much smaller numbers of
subjects, rtfMRI has similarly been devoid of significant adverse effects, even in chronic pain patients (Hawkinson et al. 2012), who are particularly susceptible to side effects. However, one study in schizophrenic patients found that they detected disgust faces more with up-regulation of anterior insula (Ruiz et al. 2011). Another risk is the maladaptive neural plasticity that could be induced, for instance, by repeated rtfMRI neurofeedback training of dysfunctional strategies. While less severe, the most common risks include mental fatigue and physical discomfort, natural accompaniments to experiments that require concentration and minimal head movement in the scanner. In addition, people may also feel claustrophobic in such a tight space, which limits the potential patient population.

3.3 Determining effect size

After showing the general applicability of rtfMRI neurofeedback for training of regulatory abilities in mental disorders (Phase 0 Trials), the method needs to undergo scrutiny by the methods of evidence-based medicine. First, determine the effects of rtfMRI neurofeedback (Jacobson and Christensen 1996) in healthy participants (Phase 1 trials). In Phase 2 trials, intended to study safety and feasibility and to develop optimal protocols ("dose-finding") in small patient groups, further aspects of efficacy encompass the development of the learned regulatory abilities after the end of the direct training period, such as follow-up examinations after 3, 6 and 12 months (e.g. (Craske et al. 1991; Öst and Westling 1995)). In rtfMRI neurofeedback, researchers have conducted a number of Phase 1 trials on healthy subjects showing behavioral effects (e.g. (Bray et al. 2007; Shibata et al. 2011; Weiskopf et al. 2003)).
Some pilot studies have reported effectiveness in patients, as required in Phase 2 trials (see Table 1), but have not reported follow-up evaluations.

3.4 Randomized controlled trials and multicenter studies

In further stages of investigation, rtfMRI neurofeedback will have to prove its clinical utility in comparison with alternative therapeutic methods, for example psychotherapy, pharmacotherapy, physiotherapy or other physical interventions. This will require randomized clinical trials (Phase 3 trials). These studies will address the efficacy and generalizability, of neurofeedback while revealing risks and side effects in comparison to other methods (Jacobson and Christensen 1996), such as biofeedback using peripheral mechanisms and experimenter guidance, likely in a multi-center design. The challenges of these Phase 3 trials include high numbers of subjects (depending on effect size), well-defined control-groups such as sham feedback and alternative methods, and close communication between participating centers to ensure data stability and consistency (e.g. multicenter studies on the efficacy and mechanisms of psychotherapy as conducted by Gloster et al. 2009).

3.5 Replication

Ensuring reproducibility is a key challenge in neurofeedback studies. One of the landmark works in the field found self-regulation of rACC could reduce pain scores in chronic pain patients compared to controls in a single session (deCharms et al. 2005). A follow-up trial consisting of six sessions, and using six different explicit mental strategies, on a larger number of subjects (21
experimental, 38 sham) showed improved ability of the experimental group to control target
ROIs compared to sham (presentation by deCharms). However, this ability did not transfer to
behavioral outcomes as previously found. Since rtfMRI neurofeedback research is relatively
expensive and time intensive, replication is an arduous task. In a recent opinion paper from the
field of "omics" (i.e. genomics, proteomics, etc.), Ioannidis and Khoury argue the necessity of
incentives to replicate results (such as funding rewards or reductions), or of setting specific
reproducibility requirements (Ioannidis and Khoury 2011). Since their research is also
expensive and time-consuming, they point to some promising solutions such as public data
repositories for larger data set analysis. It may be too early for such a solution for rtfMRI, given
the study parameters that can vary widely (see Section 1: Considerations in Study Design), and
with currently little empirical evidence showing how these changes affect neurofeedback
performance. Regardless of an eventual solution, a promising study must eventually pass the
reproducibility litmus test before much larger amounts of time, energy and money are spent on a
full clinical trial.

Caution against overstating prognoses should be exercised. Despite the proof-of-principle in
healthy subjects and preliminary results in some patient groups, a real usefulness in clinical
routine is far from being demonstrated (see open questions listed in Box 3). The reader should be
cautioned that assuming such robust results from rtfMRI is hypothetical; currently, the more
fundamental diagnostic capability of fMRI has yet to be established. Nevertheless, the growing
interest in fMRI-based neurofeedback and its clinical applications is likely to also lead to a
deeper understanding of the brain processes underlying neurological and psychiatric disorders.
4 Learning Mechanisms

While there are several studies demonstrating rtfMRI as a scientific tool or a clinical/therapeutic method, there are very few studies targeted at testing specific theoretical hypotheses about the mechanism of operant control of neural activity with feedback. Further studies in this regard, including primate experiments combining fMRI and electrophysiology could give greater insight into how BOLD regulation leads to neural changes and vice versa. Gaining an understanding of and then exploiting these learning mechanisms could help standardize and quantify methods used in the field. In this chapter, we discuss some fundamental questions from the conference regarding what model of learning best represents neurofeedback, whether learning is implicit or explicit, and which factors influence learning.

4.1 What are the learning mechanisms engaged by neurofeedback?

Literature from biofeedback and learning theory suggest two competing theories, namely, operant learning theory and cognitive-awareness view, attempting to explain learning mechanisms of volitional regulation of physiological functions in animals and humans (Black et al. 1977).

Operant learning theory states that the probability of a physiological response is increased when a reinforcing stimulus follows that response. The theory focuses on three main elements: (1) discriminative stimuli (SDs), (2) responses, and (3) reinforcers. When the response is reinforced in the presence of one SD (e.g. a visual symbol of an up-arrow) and not in the presence of other SDs (e.g. visual symbol of a down-arrow), the changes in response probability will occur only in
the presence of the first SD. In the case of a rtfMRI neurofeedback experiment, a reinforcing stimulus could be the real-time feedback of the brain activity, for example, in the form of increase in the bars of a thermometer in proportion to the amplitude of the BOLD signal in a brain region respective to baseline or a given reference activity. Although much is known regarding how reward affects learning and behavior (Schultz 2000), mechanisms of operant learning are still not completely understood.

A rival theory to operant learning theory is the “cognitive-awareness view” which states that one establishes control over a response by making a subject aware of the sensations that are produced by the occurrence of the response (Black et al. 1977). If the experimenter provides information about a response, the subject becomes aware of it, which leads to voluntary control over the response. In other words, becoming aware of a response is a necessary condition to achieve voluntary control over it. According to this theory, one may become aware of the response or the sensory changes associated with it, and that awareness itself might act as a discriminatory stimulus that the subject subsequently uses for learning to produce the response. In addition, a salient exteroceptive stimulus could also be presented to be contingent on the response as is done during a rtfMRI neurofeedback experiment.

However, there are several empirical issues with the cognitive-awareness view. Firstly, there is the problem of precisely defining what awareness means. Secondly, there is voluminous literature on learning without awareness (Dehaene and Changeux 2011; Saltz 1971). There are different factors that the subject could be aware of, including, the reinforcer, reinforcer-response contingency and the response itself. Data indicate that while awareness helps, subjects can learn
even when they are unaware of the specific response that has been conditioned, and even when
they are unaware that any response has been conditioned (Black et al. 1977). Of course it is
possible that awareness of the response might play a role in the learning of voluntary control
because interoceptive stimuli during the response could act as SDs for the occurrence of the
subsequent response. However, such awareness would not be essential because either
interoceptive or exteroceptive stimuli could act as SDs. The cognitive-awareness view asserts
that awareness is essential for voluntary control while the operant learning theory could assign
some role to awareness, but in no case claims to be neither necessary nor sufficient for
establishing voluntary control.

While there has been enormous data in support of the operant learning theory, one limitation is
that the theory does not specify the mechanism through which the change is produced. In
comparison, the awareness view appears on first look to be superior as it specifies the underlying
psychological mechanism through which response-contingent stimuli have their effect, i.e.
through awareness of the response. However, the awareness view is currently incomplete since
it is unknown how response-contingent stimuli lead to response-produced sensations, and how
awareness results in control.

4.2 What are the biological mechanisms of neurofeedback?

While much is known regarding how BOLD changes are associated with underlying neural
changes, there are still certain cases that remain unclear. It is known that both neural excitatory
and inhibitory responses lead to increase in BOLD (Logothetis 2008). Given this finding, we
cannot say whether voluntary up-regulation of BOLD leads to neural excitation or inhibition, a
process dependent on the specific brain region and function. However, one might assume that
induced neuronal activation would lead to synaptic pruning and consolidation or strengthening of
the used connections and networks.

There are at least two potential ways in which learned self-regulation of BOLD response could
modulate the underlying neural activity. One obvious mechanism is the direct neural activation
elicited by the mental imagery performed by the participant, and the variability in its intensity
and vividness over time. Changes in neural activity necessarily lead to changes in BOLD that is
then seen in the feedback signal. Another potential mechanism, although debatable, could be the
learned modulation of the vascular activity, brought about by direct instrumental control, that
may in turn influence the neural activity in a reverse fashion. Moore and Cao propose a very
controversial hypothesis (Moore and Cao 2008), suggesting that increased blood flow can
facilitate neural activity, as opposed to the canonical belief that blood flow simply represents
metabolic cost of firing neurons. Future studies of rtfMRI self-regulation, conducted during
combined BOLD and neuro-electric measurements, might help answer the above questions, not
only throwing further light on the brain mechanisms of volitional regulation, but also on neuro-
vascular coupling.

4.3 Are there limits to operant learning and how can they be overcome?

It has been shown that BCIs based on operant learning of electrical or hemodynamic brain
responses can be used by paralysed people to select letters or words with their EEG recorded
brain activity and thus restore communication (Birbaumer 2006; Birbaumer and Cohen 2007;
Birbaumer et al. 2008; Buch et al. 2012; Vaadia and Birbaumer 2009). However, despite
repeated efforts, it has thus far not been possible to train BCI-use in the completely locked-in
state and in vegetative state (Kübler and Birbaumer 2008; Ramos Murguialday et al. 2009). It
should be noted that work is currently in progress to address this specifically with rtfMRI-based spellers (presentation by Goebel, (Sorger et al. 2012)). In light of this, Birbaumer has proposed that extinction of voluntary goal-directed behavior and thinking after prolonged periods of complete lack of movement contingencies is responsible for this failure despite intact input-oriented cognitive processing. Birbaumer pointed out that the problem of replicating operant learning of autonomic responses in the curarized rat (Miller 1975; Miller and Dworkin 1974) may constitute an analogue to the failure of brain communication in complete paralysis. Technical alternatives to non-invasive BCIs such as communication with saliva pH-changes (Wilhelm et al. 2006), sniffing (Plotkin et al. 2010), functional near infrared spectroscopy (fNIRS; (Naito et al. 2007; Sitaram et al. 2007) and invasive recordings (Hochberg et al. 2006; Ramos Murguialday et al. 2009) cannot overcome the described psychological learning deficit which may be even more profound in vegetative ("apallic") state where patients spend years in unresponsive positions despite partially intact cognition (Kotchoubey et al. 2005; Monti et al. 2010). The above problem could reflect a physiological and methodological constraint to any form of operant training including rtfMRI neurofeedback. In view of this problem, Birbaumer and colleagues have planned to experimentally test a solution to this fundamental disorder of volition and loss of communication abilities through two-process learning of brain-responses: first by using “reflexive” classical conditioning of brain responses, and second by the process of acquisition paired with reward, through instrumental conditioning (Liberati and Birbaumer 2012; Liberati et al. 2012).

4.4 Is rtfMRI neurofeedback implicit or explicit learning?
Many neuroscientists and psychologists not in the mainstream of learning theory continue to hold cognitive (explicit) and conditioning (implicit) theories as two separate and opponent concepts (Song et al. 2012). Kirsch and colleagues reviewed data pertaining to the role of higher-order cognition in conditioning and proposed a theoretical synthesis of both automatic and cognitively mediated processes (Kirsch et al. 2004).

Contemporary mechanistic accounts of conditioning are based on the hypothesis that direct Stimulus-Outcome (S-O) and Response-Outcome (R-O) associations are formed during learning. However, there is ongoing disagreement whether these are simple associations at the level of cognition in these processes. Higher-order cognitive theories are centered on the concept of expectancy, defined as a future-oriented belief, as more than the activation of simple binary associations. Instrumental conditioning often produces a belief that particular behaviors will produce particular outcomes (e.g. moving a certain lever brings food). A large body of data, especially from Tolman (Tolman and Minuim 1942; Tolman 1948) indicated that rats in mazes behaved as if they had access to information, built cognitive maps of the mazes and expected to find food in particular locations (Bolles 1979). While traditional conditioning literature predominantly supports mechanistic theory, there is abundant new data suggesting cognitive interpretation in some situations. Recently, Kirsch et al. (Kirsch et al. 2004) proposed a less parochial interpretation that some conditioning may happen with cognition while others without it, and that there are other learning processes that may not use conditioning at all (e.g. learning by observation or verbal communication). In higher organisms like humans, behavioral flexibility requires greater complexity, forming associations via conditioning procedures as well as from other sources of information. Kirsch et al. speculated that the more complex the
This above discussion is relevant to the questions raised in the conference as to whether rtfMRI training should involve explicit or implicit mental strategies for learning self-regulation, and which of the two is superior for learning. In this light, it appears that learning may involve both methods, leading one to anticipate that the role of cognition will play in the learning process as the difficulty of brain regulation increases, a hypothesis that could be investigated with rtfMRI methods (Dayan and Cohen 2011).

4.5 What are the factors that influence learning?

Although there is a large body of literature concerning learning and conditioning, it is astonishing that relatively little of theoretical and experimental concepts of operant conditioning have been mobilized in neurofeedback research (Schwartz and Beatty 1977). However, biofeedback literature and recent findings from rtfMRI studies suggest the following as the major factors that affect learning and that can be manipulated: contingency, contiguity and time delay, instruction reinforcement, and manipulations such as shaping and chaining.

4.5.1 Contingency refers to the conditional probability of reinforcement given a response and given a failure to respond. The study of this factor includes the investigation of different modalities (visual, auditory, tactile etc.) of the response-contingent stimuli, their different physical properties (such as amplitude, rate and the complexity), and the different functional
relationships between the response and the feedback. The rtfMRI research community can learn much from the EEG biofeedback literature. For example, in studies of alpha density conditioning in human subjects, visual and auditory feedback stimuli indicated that visual feedback was more beneficial for learning (Lynch et al. 1974; Travis et al. 1975). In addition, comparisons have been made in human participants between binary and proportional or continuous feedback with varying results (e.g. (Blanchard and Young 1974)).

4.5.2 Temporal contiguity refers to the time interval between response and reinforcement. In an rtfMRI setting, the intrinsic delay between the neural activity that is regulated and the BOLD changes due to the tardiness of the hemodynamic response is known. Added to this would be the delay in acquisition and computation of the feedback signal. In general, one may assume that the continuous feedback will have advantages over intermittent feedback in terms of maintaining greater attention, providing better contiguity between response and reinforcement. On the other hand, it may have disadvantages, such as introducing cognitive and working memory overloads, and task conflict, especially when subjects employ mental imagery to self-regulate brain regions.

Johnson et al. compared between continuous and terminal (intermittent or blocked) feedback in 13 participants who performed a motor imagery task to increase the BOLD levels in the left premotor cortex with rtfMRI feedback using a cross-over design (Johnson et al. 2012). Their results indicate that intermittent presentation of feedback (about 20 s delay) is more effective than continuous presentation when an imagery-based strategy was used for self-regulation. In line with the discussion above, intermittent feedback could be more advantageous because it reduces noise, is not dependent on the hemodynamic delay, and does not interfere with imagery during self-regulation.
Certain MR signal acquisition and processing related issues, such as signal to noise ratio (SNR), spatial resolution, temporal resolution, and image spikes and artifacts, would conceivably influence contingency and contiguity of the feedback, adversely affecting learning when there is a drop in their quality and reliability. While currently only unpublished data investigates the discriminability of the feedback signal from noise (presentation by deCharms), it was suggested that a more thorough investigation of optimal SNR should be conducted in pilot testing. Future work should conduct a thorough investigation of the above issues.

4.5.3 The experimenter can generate feedback effects by varying the time delay between subject’s response and the stimulus. These effects are usually undesirable, but, for the experimenter, they provide a convincing demonstration that a feedforward relationship must exist (Miall et al. 1993; Mulholland 1977). In EEG biofeedback, there is an intrinsic time delay between the alpha response and the occurrence of the effects of a stimulus of about 250 milliseconds. There are further latencies between the occurrence of the stimulus and the attenuation of alpha waves. Delayed feedback has been shown to cause oscillations between the alpha waves and the feedback in the form of an on-off mode (Mulholland and Peper 1971). By varying the time-delay in the feedback one can study its effect on the variables such as the brain response and behavior, showing convincing evidence for a causal path existing between the independent and dependent variables.

4.5.4 Methods employed for training voluntary control generally contain two procedural elements: instructions and response-contingent stimulation. Although for reasons of historical
bias, the experimental analysis of voluntary control has tended to emphasize the role of feedback and has neglected the influence of instructions, data suggests that instructions are not at all neutral in influencing voluntary control. Instructions not only influence what the subjects will do in the experimental situation, but also what they will say they are doing or experiencing. Subjects tend to report pleasant experiences if the instructions lead them to expect such experiences. In view of this, the investigation of experimental instructions is of fundamental importance to the analysis of voluntary control. Furthermore, psychological and behavioral tasks designed to test the effect of neurofeedback training must look for sensitive yet robust measures that control for effects of placebo and instruction.

While the majority of rtfMRI studies to date have instructed subjects to use mental imagery to increase or decrease the BOLD response in circumscribed brain regions (see references in Introduction), only two recent studies (Kim et al. 2011; Shibata et al. 2011) have reported direct operant training occurring without the use of explicit mental imagery (See Section 1: Considerations in Study Design for debate on explicit and implicit instructions).

4.5.5 Shaping and chaining have received attention in the biofeedback literature (Black et al. 1977) but have not been explicitly elaborated and investigated in rtfMRI studies. Shaping is a way of adding new responses or behavior to a human’s or animal’s repertoire. In shaping, the subject is first rewarded for an approximation of the target response. Successive approximations are reinforced until the target response is reached. In rtfMRI neurofeedback, subjects could be trained to produce a spatial pattern of brain activity by successive approximations of the target pattern. Chaining is a series of responses or behaviors where one behavior is a cue to the next
and a chain of these responses lead to a complex behavior. In rtfMRI neurofeedback, subjects
could be trained to produce a temporally extended set of distinct brain responses by chaining one
response to another in a series. Subjects may first increase the BOLD response in a single region
at a given time-point, followed by an increase of BOLD in a second region at the next time-
point, followed in the end by the decrease of BOLD in a third region in the following time-point.

4.6 Time-dependence of Learning

Learning occurs at different time-scales (Doyon and Benali 2005), which have distinguishable
neural correlates (Floyer-Lea and Matthews 2005; Shadmehr and Holcomb 1997). Over time,
learned abilities are consolidated and then transferred to long-term memory. Reconsolidation is a
phenomenon in which recall of learned content initiates a second consolidation process
(presentation by Cohen). Neurofeedback studies have shown effects of retention over multiple
sessions, suggesting both consolidative and reconsolidative processes were at work (Bray et al.
2007; Shibata et al. 2011). These particular studies used monetary reward to assist learning the
neurofeedback signal; in agreement with the results of Abe et al. which found improved long-
term learning of a motor task with monetary reward (Abe et al. 2011). However, neurofeedback
studies have not examined the long-term learning effect.

It appears most likely that operant mechanisms underlie learning in rtfMRI neurofeedback.
However, operant learning may have limitations that could prevent use by locked-in patients who
are unable to elicit a voluntary command. The nature of the feedback itself can be deconstructed
into components that may individually or in concert affect learning. How these parameters may
be tuned and to what extent remains one of many open questions (see Box 4).
Where is the future of rtfMRI neurofeedback?

Since its introduction in 1995 (Cox et al. 1995) rtfMRI has inspired research leading towards neural intervention, intraoperative procedures, brain-computer interfaces and quality assurance. While the future of rtfMRI neurofeedback can lead towards some exciting applications in a multitude of neurological disorders, we are currently just beginning to scratch the surface of where it can be applied. This section discusses both the immediate and long-term future of fMRI-based neurofeedback.

5.1 Immediate Future

Naturally, the future of rtfMRI coincides with that of fMRI (for review on advances in fMRI, see (Wald 2012)). Recent work has made measurement of more specific regions possible. For instance, imaging of the function of microcolumnar structures is being implemented, including using higher static fields such as 7T and high resolution grid sampling (presented by Goebel). Goebel mentioned the availability of ultrafast sequences that could allow very low TRs, improving contrast-to-noise ratio. This strategy is already being investigated by other groups (Posse et al. 2012). Another method of obtaining more specificity using MVPA was presented, distinguishing cortical representation of individual fingers in real time in the primary somatosensory cortex (presented by Kaas). Taken together, fMRI is becoming more specific and faster.
Hyperscanning is a technique developed to measure brain activity during social interaction (Montague et al. 2002), which can be combined with neurofeedback training (Goebel et al. 2004b). A successful implementation of rtfMRI hyperscanning was presented using navigation through simple competitive and cooperative tasks through motor imagery, and a further example implementing a virtual environment to examine cooperation (presentation by Baecke). Hyperscanning, especially in a virtual environment, has potential for use in social neuroscience experiments, specifically neuroeconomics studies, human-computer interaction and human-computer-human interaction.

Magnetic resonance spectroscopy (MRS) can quantify the concentration of certain specific chemical compounds and has the ability to measure neural correlates of neurotransmitters, for example GABA and creatine (Castillo et al. 1996). To date, real-time functional MRS has already been used to quantify dynamic BOLD changes in real-time (Koush et al. 2011). In the future, real time MRS could be used to manipulate neurotransmitter production or track brain metabolites (presentation by Koush).

Another alternative to traditional fMRI is arterial spin labeling (ASL). ASL traces arterial blood as it flows into the brain (regional cerebral blood flow, rCBF) by “tagging” arterial blood magnetically and then measuring the response approximately one second later in the brain (Detre et al. 1994) and comparing it to a "non-tagged" control condition. It has poorer temporal resolution than EPI as a result but the advantages of a physiological and clinically meaningful outcome measure in rCBF and a true baseline. Real-time ASL (rtASL) has recently been reported (Hernandez-Garcia et al. 2011). At the conference, researchers at the University of
Tuebingen reported results from rtASL using a surround-subtraction method to calculate the feedback signal in approximately three seconds (presentation by Várkuti). Although currently the signal-to-noise ratio is not as good as EPI and there are still issues with selecting the optimal feedback region, the future for rtASL is promising due to its inherent advantages over BOLD signal related methods. Some potential uses of rtASL include tracking thrombolysis in ischemic stroke or anesthesia depth. Indeed, the disadvantage of the lower temporal resolution could be negated through experiments focused on brain regions that cannot be so quickly modulated, since ASL does not suffer from signal baseline drifts like BOLD imaging.

Feedback of network or connectivity-related activity may better represent brain physiology than region-based methods and exploit the advantages of whole brain coverage. Some trends towards MVPA in rtfMRI were presented at the conference (presentations by LaConte and Goebel), including an experiment examining SVM classification of emotional states (presentation by Rana), as well as recently published examples using sparse logistic regression (Shibata et al. 2011). Some examples of connectivity feedback included experiments using connectivity between the inferior frontal and superior temporal gyrus (presentation by Ruiz) and between bilateral motor cortices (presentation by Zilverstand). Both functional and effective connectivity methods, as well as multivariate pattern classification, could represent part of a larger wave of movement towards multivariate feedback.

5.2 The Longer Term Future
Predicting where such a fast-changing field will emerge in the next two decades is a difficult task. The current direction would suggest higher contrast-to-noise, more physiologically-related signals from multiple, more precise areas of the brain will be accessible. At the same time, more advanced computational methods of state classification and signal conditioning are being developed that will further improve robustness and selectivity of rtfMRI. As a result, rtfMRI protocols will likely become more varied before they begin to settle to some accepted design principles.

For therapeutic purposes, one would assume that clear physiological signals facilitate better neurofeedback performance. Subsequently, the functional consequences of such self-control will become more clearly defined, and thus more accurately identify ideal candidates. Standardization of transfer will be established and compared with specific behavioral and psychological measures during neurofeedback. It is well imaginable that in the next decades rtfMRI neurofeedback could enter the phase of clinical treatment of specific neurological or mental disorders where invasive intervention is not appropriate. Neurofeedback could also be used as a complement with other therapeutic methods, e.g. physical rehabilitation delivered via MR-compatible robotic manipulation (Gassert et al. 2008; Gassert et al. 2006). While training in the scanner may not be feasible on a long-term basis, the aim would be to have the patient transfer this learned ability for use ubiquitously outside the clinic.

Apart from use of rtfMRI neurofeedback for therapy, the application of rtfMRI in psychiatry could also consist in identifying the neural correlates of certain mental or psychotherapeutic interventions in patients. Such diagnostic procedures could help prove the effectiveness of
psychotherapeutic sessions or identify eligible patients for certain psychotherapies, specifically towards treatment response prediction. It may also serve for gaining knowledge about the neurobiological backgrounds of mental interventions applied in a psychotherapeutic context.

FMRI BCIs could be useful in applications that require precise measurement of whole brain activity. This technology could be of great use for patients unable to communicate by any other means, including EEG-based BCIs (Sorger et al. 2012). For bedside BCIs used daily, expensive, stationary and slow technology such as fMRI may not be feasible, but for situations that require relatively high spatial resolution compared to EEG, fNIRS could lead the way (presentation by Zimmerman).

Perhaps the most immediate application of rtfMRI is in quality assurance (Weiskopf et al. 2007). As scanner manufacturers further implement real-time packages in their own software, clinicians will be able to ensure contrast integrity, motion parameters and identify electromagnetic interference with additional research or clinical equipment. In the future, it may be possible that scanner sequences will optimize themselves to improve contrast, and maybe even adjust for movement artifacts. The ability to compensate for movement online, perhaps using ultrafast sequences, would open up a whole new range of tests that could be conducted in the scanner.

There are still fundamental questions about rtfMRI that may need to be addressed before the technology is ready for clinical translation. Issues such as learning to control the signal with the hemodynamic delay, optimal sensory channels for feedback and feedback design, how to maximize contrast-to-noise ratio, whether the effect size is clinically relevant, and whether the
training can be transferred outside the scanner are critical to understand. While many studies
have addressed some of these issues within a specific application, these questions should be
revisited before applying rtfMRI to any new field. In particular, what are the best strategies to
facilitate learning to control brain activity across modalities (Censor et al. 2012).

6 Conclusions from the meeting

Over the past decade much work has shown promise for rtfMRI in neurofeedback and other
applications. Some key successes, including showing relevant behavioral effects of
neurofeedback, exhibiting its use as a scientific tool, and identifying online brain states have led
to a recent spike in interest in the field. Yet despite clear progress, fundamental issues remain
such as the minimum discernible signal-to-noise ratio of feedback, causal experimental designs,
imagery strategy, effect size, transfer, and how participants learn to self-regulate their BOLD
signal. Following the talks and discussions, it was agreed that these issues and current ones
should be discussed every two years at this conference, open to the worldwide community. In
addition, a mailing list was created to share general thoughts, problems, job announcements, or
other relevant information to the field (email James Sulzer at jamessulzer@gmail.com to
register).

Acknowledgements

The authors would like to thank the sponsors of the event, including the Swiss National Science
Foundation (project #31CO30_139955), the Zurich Neuroscience Center (ZNZ), Philips
Medical, ETHZ, and the University of Zurich.
**Figure Captions and Tables**

**Figure 1:** Results of literature search regarding rtfMRI, found using the search term "Real-time fMRI" in Google Scholar, restricting findings to journal publications that use rtfMRI, develop technology specifically for rtfMRI or reviews primarily about rtfMRI. Publications were then categorized and color-coded accordingly.

**Figure 2:** Schematic of rtfMRI control loop. Typically, EPI images are extracted from the MR scanner online, analyzed by third-party software, and then presented back to the subject for the purposes of neural self-regulation. Reprinted from (Weiskopf et al. 2004b) with permission from Elsevier.

**Table 1: Overview of studies using real-time neurofeedback in patients suffering from various neurological and psychiatric disorders**

Control subjects generally received no feedback or no real feedback (“sham-feedback”). *healthy subjects as control participants, furthermore other groups receiving different forms of feedback and training (4 patients, 24 healthy subjects), **healthy subjects as control group

<table>
<thead>
<tr>
<th>Study</th>
<th>Disorder</th>
<th>N subjects/ control group</th>
<th>Brain regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(deCharms et al. 2005)</td>
<td>Chronic pain</td>
<td>12/36*</td>
<td>ACC</td>
</tr>
<tr>
<td>(Ruiz et al. 2011)</td>
<td>Schizophrenia</td>
<td>9/0</td>
<td>Insular cortex</td>
</tr>
<tr>
<td>(Haller et al. 2010)</td>
<td>Chronic tinnitus</td>
<td>6/0</td>
<td>Auditory cortex</td>
</tr>
<tr>
<td>(Subramanian et al. 2011)</td>
<td>Parkinson’s disease</td>
<td>5/5</td>
<td>Supplementary motor complex</td>
</tr>
<tr>
<td>(Linden et al. 2012)</td>
<td>Major depression</td>
<td>8/8</td>
<td>Brain regions involved in positive emotions (VLPFC R/L, insula r cortex R/L, DLPFC)</td>
</tr>
</tbody>
</table>
(Sitaram et al. 2012)  Chronic stroke  2/4**  Ventral premotor cortex L

Abbreviations: ACC anterior cingulate cortex, VLPFC ventrolateral prefrontal cortex, DLPFC dorsolateral prefrontal cortex, OFC orbitofrontal cortex, R right, L left

Table 2: Comparison of tools for neural interventional tools used in scientific investigation in humans (rated from +++ = most advantageous to - - - = least advantageous). Relative ratings based on experience and opinions of the authors.

<table>
<thead>
<tr>
<th></th>
<th>Independence</th>
<th>Specificity</th>
<th>Repeatability</th>
<th>Controllability</th>
<th>Multiple regions</th>
<th>Invasiveness</th>
<th>Time Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMS</td>
<td>++</td>
<td>+</td>
<td>+ ++</td>
<td>+ +</td>
<td>-</td>
<td>+</td>
<td>+ +</td>
</tr>
<tr>
<td>tDCS</td>
<td>+ +</td>
<td>- -</td>
<td>+ ++</td>
<td>+ +</td>
<td>-</td>
<td>+</td>
<td>+ +</td>
</tr>
<tr>
<td>DBS</td>
<td>+ + +</td>
<td>+ + +</td>
<td>+++</td>
<td>+ +</td>
<td>-</td>
<td>- -</td>
<td>+ + +</td>
</tr>
<tr>
<td>EEG</td>
<td>+</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>+ +</td>
<td>+ + +</td>
<td>+ + +</td>
</tr>
<tr>
<td>Focal lesions</td>
<td>+ + +</td>
<td>++</td>
<td>- -</td>
<td>- -</td>
<td>+</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>Neuropharmacology</td>
<td>+ + +</td>
<td>- -</td>
<td>+ +</td>
<td>+ +</td>
<td>-</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>rtfMRI</td>
<td>+</td>
<td>++</td>
<td>- -</td>
<td>- -</td>
<td>+ +</td>
<td>+ + +</td>
<td>+</td>
</tr>
</tbody>
</table>
Box 1: Open Questions for Study Design

1. What is the optimal study design (i.e. run length, block length, etc.) for learning and how does it vary with region?
2. Should instructions be implicit or explicit?
3. How does the research goal affect the choice of control groups?
4. Are visual feedback channels always more advantageous than other sensory channels?
5. How much information should be conveyed in feedback?
6. When is intermittent feedback more advantageous than continuous feedback?
7. Do the advantages of differential ROI feedback (i.e. cancels out unspecific effects and provides a within-subject control) outweigh the disadvantages (i.e. increased noise)?

Box 2: Open Questions for Scientific Applications

1. Can it be proven that excitation, endogenously or exogenously elicited, truly causes an action?
2. Should mental strategy choice be limited in causal rtfMRI experiments?
3. How can controllability, repeatability, specificity and independence in rtfMRI neurofeedback be improved?
4. What control experiments need to be run in order to establish specificity/causal links?
5. How can introspective measures related to mental strategies be quantified/classified (participant debriefing)?

Box 3: Open Questions for Clinical Applications

1. In which neurological diseases is rtfMRI neurofeedback appropriate, and under what conditions is it inappropriate?
2. Under which conditions is rtfMRI neurofeedback more advantageous than other interventions?
3. To what extent is the behavior of healthy participants a model for patients?
4. Can self-regulation be repeated outside the clinic?
5. How effective is the treatment?
6. What are the side-effects?
7. How long does the training effect last?

Box 4: Open Questions for Learning Mechanisms

1. What is the optimal study design (i.e. run length, block length, etc.) for learning and how does it vary with region?
Appendix: List of Presenters (alphabetical order)

Sebastian Baecke (Otto-von-Guericke University in Magdeburg)
Niels Birbaumer (University of Tuebingen, Oespedale San Camilo, Venice)
Maria Laura Blefari (ETHZ)
Annette Bruehl (Psychiatric University Hospital Zurich)
Leonardo Cohen (National Institutes of Health)
Christopher deCharms (Omneuron)
Rainer Goebel (University of Maastricht)
Sven Haller (University of Geneva)
Maurice Hollman (Max Planck Institute Leipzig)
Amanda Kaas (Maastricht University)
Yury Koush (Aachen University)
Stephen LaConte (Virginia Tech University)
David Linden (University of Bangor)
Mohit Rana (University of Tuebingen)
Sergio Ruiz (University of Tuebingen)
Frank Scharnowski (Psychiatric University Hospital Zurich)
Ranganatha Sitaram (University of Tuebingen, University of Florida)
James Sulzer (ETHZ)
 Bálint Várkuti (University of Tuebingen)
Nikolaus Weiskopf (University College London)
Anna Zilverstand (University of Maastricht)
Raphael Zimmerman (ETHZ)

1. To what extent do feedback factors such as feedback delay, contingency, reinforcement, motivation, instructions and manipulations differentially affect learning?
2. What neural correlates underlie neurofeedback learning?
3. How do explicit and implicit strategies affect learning?
4. Can participants create new cognitive strategies for improving performance, or do they simply cycle through existing ones?
5. Is the level of activation using neurofeedback that can be reached greater than that with a predefined task?
6. Does state of the feedback signal is discernible to the subject, and to what extent can the signal be discerned spatially and temporally?


