

# Action and Language Mechanisms in the Brain: Data, Models and Neuroinformatics

**Michael A. Arbib, James J. Bonaiuto, Ina Bornkessel-Schlesewsky, David Kemmerer, Brian MacWhinney, Finn Årup Nielsen, et al.**

**Neuroinformatics**

ISSN 1539-2791

Neuroinform

DOI 10.1007/s12021-013-9210-5



**Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# Action and Language Mechanisms in the Brain: Data, Models and Neuroinformatics

Michael A. Arbib · James J. Bonaiuto ·  
Ina Bornkessel-Schlesewsky · David Kemmerer ·  
Brian MacWhinney · Finn Årup Nielsen · Erhan Oztop

© Springer Science+Business Media New York 2013

**Abstract** We assess the challenges of studying action and language mechanisms in the brain, both singly and in relation to each other to provide a novel perspective on neuroinformatics, integrating the development of databases for encoding – separately or together – neurocomputational models and empirical data that serve systems and cognitive neuroscience.

**Keywords** Linking models and experiments · Models, neurocomputational · Action and the brain · Language and the brain · Mirror systems · Multi-level data · Multi-level models ·

M. A. Arbib (✉)  
Computer Science and Neuroscience Graduate Program, University  
of Southern California, Los Angeles, CA, USA  
e-mail: arbib@usc.edu

J. J. Bonaiuto  
Division of Biology, California Institute of Technology, Pasadena,  
CA, USA  
e-mail: jbonaiuto@gmail.com

I. Bornkessel-Schlesewsky  
Neurolinguistics, University of Marburg, Marburg, Germany  
e-mail: iboke@staff.uni-marburg.de

D. Kemmerer  
Speech, Language, & Hearing Sciences and Psychological Sciences,  
Purdue University, West Lafayette, IN, USA  
e-mail: kemmerer@purdue.edu

B. MacWhinney  
Psychology, Computational Linguistics, and Modern Languages,  
Carnegie Mellon University, Pittsburgh, PA, USA  
e-mail: macw@cmu.edu

F. Å. Nielsen  
Technical University of Denmark, Copenhagen, Denmark  
e-mail: fn@imm.dtu.dk

E. Oztop  
Ozyegin University, Istanbul, Turkey  
e-mail: Erhan.Oztop@ozyegin.edu.tr

Databasing empirical data · Federation of databases ·  
Collaboratory workspaces

## Bridging the Gap Between Models and Experiments

The present article offers a perspective on the neuroinformatics challenges of linking neuroscience data with models of the neural and other interactions that yield those data. We will focus on cognitive and systems neuroscience. The perspective is particularly informed by consideration of the brain mechanisms underlying action and language and emerged from discussions held as part of the Workshop on “Action, Language and Neuroinformatics” held in July of 2011 (see the section “Acknowledgement, Discussion Groups and Workshop Participants” for further details). The [Appendix](#) provides a Tabulation of Present and Future Resources relevant to the issues discussed in this article.

## What is a Model?

For one researcher, a model may be a technique for marshaling diverse data into a coherent format, whereas for others a model (conceptual or computational) provides an account of how interactions of entities within the brain mediate between inputs, internal states and outputs. We refer to the former as “data models” and the latter as “processing models.” Clearly, the result of data modeling is crucial to any specification of what it is that a processing model has to explain. If the word “model” is used in what follows without a qualifying adjective or context, it will be in the sense of a “processing model”.

There is a third type of model – not a model of data, not a model *of* the brain, but a model *in* the brain. This idea goes back, at least, to Craik (1943). It relates both to the general notion of perceptual schemas and motor schemas (Arbib 1981), the control theory concepts of feedback and

feedforward, and the notion of forward and inverse models of a system (Wolpert and Kawato 1998). Thus, in modeling neural mechanisms for action and action recognition, processing models of how the brain employs forward and inverse models may play a crucial role (Oztop et al. 2013; Oztop et al. 2005).

### Fostering Modeler-Experimentalist Collaboration

Even among experimentalists who have rich interactions with modelers, few make explicit what challenges – whether at the level of explaining specific data or in search of a conceptual framework – they want modelers to address, and few will adjust their agenda to test novel predictions of models. Thus a challenge of particular interest here is to chart how neuroinformatics could provide means to deepen these interactions. Among the relevant issues are understanding how to summarize data sets into a form for which modeling is appropriate, and appreciating the value of models which do not fit data but do provide fresh insights – this in addition to the more obvious appreciation of those which do so – while avoiding something like the “epicycles” used to adjust the orbits of Ptolemaic astronomy, i.e., without introducing ad hoc mechanisms whose only *raison d'être* is to explain a very limited data set.

Arbib and colleagues (Arbib et al. 2014b) argue for indexing models not only with respect to brain structures (e.g., a model of circuits in basal ganglia and prefrontal cortex) but also with respect to brain operating principles (BOPs) which provide general mechanisms (such as reinforcement learning, winner-take-all, feedforward-feedback coupling, etc.) which may be employed in analyzing the roles of very different brain regions in diverse behaviors. Moreover, they argue that each model should be associated with summaries of empirical data (SEDs) defined at the granularity of the model. There are at least two problems here. (a) Even if experimentalists make clear the exact methodology used to extract data and process it – as in the framework offered by Lohrey et al. (2009) to integrate an object model, research methods (workflows), the capture of experimental data sets and the provenance of those data sets for fMRI – the problem remains of integrating data gathered with different protocols into a meaningful challenge for modeling. (b) If model A explains view A of data set D, while model B explains view B of D; and yet the models are different, how may one build on them to more fully address aspects of D revealed in the combination of the 2 views? For example, one model might be successful if it can explain the averaged responses of a brain region to a key set of stimuli rather than explaining individual variations, whereas another model might be designed to explain key patterns of individual variation (e.g., aphasic versus non-aphasic).

Arbib et al. introduce the Brain Operation Database (BODB, <http://bodb.usc.edu/bodb/>) as a particular implementation of

this general framework. BODB requires that SEDs associated with a model be divided (at least) into those which are used to design the model, and those used to test the model. For computational models, the latter SEDs are compared with summaries of simulation results (SSRs) obtained using the model.

If there are summary data, then somewhere there are unsummarized data. Often such data are only available, if at all, in Supplementary Material for a published paper, or on a laboratory Website. This raises two challenges: The development of further Websites for the integration of unsummarized data and the linkage of summarized data (perhaps in another Website) to the data they summarize. For example, the huge volumes of data associated with each individual fMRI scan in one comparison for a specific experimental-control condition comparison may be summarized into brain imaging tables which might aggregate multiple scans for that comparison into a single table, losing detail but, hopefully, gaining conceptual clarity in the process. BrainMap (Laird et al. 2005), [brainmap.org](http://brainmap.org), provides the classic repository for such brain imaging data, and Laird et al. (2009) discuss the potential analyses that are possible using the BrainMap database and coordinate-based ALE (activation likelihood estimation) meta-analyses, along with some examples of how these tools can be applied to create a probabilistic atlas and an ontological system describing function-structure correspondences. Nielsen (2013) focuses on neuroinformatics tools of the BrainMap-based Brede Wiki, <http://neuro.imm.dtu.dk/wiki/>, which support federation and interaction of brain imaging data. (See the section on “Federation” below for more on the general issues concerning federation.).

A range of brain models has been documented in the ModelDB database, <http://senselab.med.yale.edu/modeldb/>. Model DB and BODB seem good targets for federation. The former provides documented code for each model, with instructions on how to run it to get published results; the latter links each model to the summaries of experimental data used to design and test it, and provides tools to retrieve and compare models which address related sets of data. While BODB tends to focus on systems and cognitive neuroscience models “up a level” from many of the models in ModelDB, Bohland et al. (2013) have focused on the challenges of linking data at multiple levels from genes and gene expression that guide brain development and provide the molecular signatures of populations of neurons to the architecture of functional brain subsystems. How can modelers begin to build systems that span the layers between information processing in brain circuits – the typical domain of neural processing models – and genes, gene products, and regulatory networks that may be at the root of disorders involving abnormal development of skilled action and language? How can we link the Allen Brain Atlas (Jones et al. 2009) ([www.brain-map.org](http://www.brain-map.org))

database of gene expression patterns in the mouse brain to emerging databases for monkey, human and other species while addressing neural homologies? We may note earlier efforts associated with the NeuroHomology Database, NHDB, to look at homologies between macaque and human brains relevant to the linkage of macaque neurophysiology to human brain imaging (Arbib and Bota 2003; Bota and Arbib 2004), but, unfortunately, NHDB is no longer available. These efforts included algorithms designed to address the problem of linking neuroanatomical data both across atlases within a species and across species (homology). A number of ideas from NHDB are incorporated in the Brain Architecture Management System, BAMS (<http://brancusi1.usc.edu/>) which supplies inference engines and five interrelated modules (Bota and Swanson 2008): Brain Parts, Cell Types, Molecules, Connections (between regions and cell types), and Relations (between parts identified in different neuroanatomical atlases).

### Action, Language and the Brain: A Brief Background

We first review data and models linking mirror neurons (in monkeys) to mirror systems (in humans) and their implications for mechanisms for action recognition, imitation, and language and their evolution. We then turn to the specific challenges for model building and database construction and federation posed by neurolinguistics.

#### Mirror Neurons and Mirror Systems as a Bridge from Action to Language

In the 1990s it was discovered that some of the neurons in an area of premotor cortex of the macaque brain dubbed F5, whose firing was specific for the monkey's grasping action (e.g., precision grip but not power grip) were also active when the monkey observed a human or other monkey performing a similar action (di Pellegrino et al. 1992; Gallese et al. 1996). These were dubbed mirror neurons, with canonical neurons being those active when the monkey executes the grasp but not when he observes it. More generally, then, a mirror neuron for an action is one which is active both when the animal executes a specific action and when it observes more or less similar actions executed by another. Since there are few single neuron recordings in humans, the interest in mirror neurons led to brain imaging studies which sought and found *mirror systems* in the human brain in the sense of regions that show relatively high activation in PET or fMRI studies both when the subject executes any of a broad set of actions or observes execution of actions from that class (see Rizzolatti and Sinigaglia 2008 for a comprehensive overview). A mirror system for grasping was found in the human brain in or near Broca's area, traditionally thought of as a speech area but

better thought of as a language area since it is involved in sign language as well as spoken language (Poizner et al. 1987). This fueled the so-called *mirror system hypothesis* that mirror neurons for grasping evolved into mirror systems for communication via manual gesture as a major step toward the evolution of the human language-ready brain (Arbib 2012; Rizzolatti and Arbib 1998). It should be noted that monkeys have mirror neurons but very little capacity for imitation, but it is generally agreed that a mirror system is part of the ape and human capacity for imitation: Imitation requires augmenting a capacity to recognize an action already in one's own repertoire with the ability to incorporate a novel observed action performed by another into one's repertoire.

Demiris et al. (2013) study information processing in the mirror neuron system. Models of the mirror neuron system in monkeys both inform study of human mirror systems and have been used for enabling robots to imitate and learn how to perform tasks from human demonstrations. The paper uses BODB to address the challenges to neuroinformatics following from the diversity of experimental setups, methodologies, and computational structures when, for example, we seek to interpret imitation neuroimaging results for humans in terms of neurophysiological recording of mirror neurons in the macaque. Moreover, the definition of "mirror system" remains fluid. Aziz-Zadeh et al. (2011) report that congenital amputees (born with no forearms or hands) nonetheless have a "mirror system" for grasping actions. They propose that this system contains "visually learned mirror neurons" – but, of course, such neurons could not be mirror neurons in the strict sense of firing for both execution and observation. However, it makes sense to posit *potential* mirror neurons which have the potential to become mirror neurons *but need not do so*. Since humans can learn by observation, we need to form a visual image of an action before we can acquire the action for ourselves, making it possible for us (like congenital amputees) to get neurons tuned only "halfway" to becoming mirror neurons when we do not go on to acquire the ability to execute the action (Arbib 2012). Future computer modeling of imitation in mirror neurons would make this clear, showing potential mirror neurons at varied stages of exploiting their potential.

Surprisingly, the functional role of mirror neurons in monkeys has not been established – the evidence shows that they exist in the sense of being active both for execution and observation of similar actions, but there are few data recorded from monkeys during social interaction to assess the actual role of mirror neurons in social behavior (see, e.g., Fujii et al. 2008; Santos et al. 2011). This points to the need for at least two (linked) repositories for neurophysiological data: one of neurophysiological data linked to behavior, with the data organized so that one may extract summaries showing parametric variation across similar tasks under similar conditions; and the other linking neurophysiological data to

dyadic interactions, since for *social* neuroscience to advance, ways must be found to track neural activity in conspecifics of different species as they interact with one another. Compare the call for a comparative neuroprimatology advanced by Gasser et al. (2013). It should also be stressed that although it is assumed that regions active in human brain imaging studies for both observation and execution of a class of actions (human mirror *systems*) depend on mirror neurons, there is (almost) no evidence for mirror *neurons* in human. More disturbingly, many imaging studies showing activity in these regions impute these activity to mirror neurons even when “mirroring” is not involved – unfortunate since we know that in monkey mirror neurons are associated with other cell types. Meanwhile, new data are emerging for comparison of functionally related neural structures in monkey and human (Orban et al. 2004; Peeters et al. 2009), with the same experimental fMRI paradigm employed in the two species (Vanduffel et al. 2002). During most of these experiments the human and monkey subjects stare at the same visual stimuli (e.g., Nelissen et al. 2005). Until recently, comparative fMRI studies relied on spatial assumptions related to cortical expansion models to bring the data from the two species in register. Typically, putative homologous areas were used as landmarks in monkeys and humans to align the fMRI activation maps, and to identify inter-species functional similarities (the Caret approach; as used in, e.g., Denys et al. 2004). However, the premise that homologous areas are both anatomically and functionally equivalent cannot be taken for granted. Mantini et al. (2012) therefore developed the inter-species activity correlation method which compares activations in the temporal domain, and which is devoid of any spatial assumption. This method may be crucial for defining cortical regions that correspond functionally but not anatomically, and will thus be relevant to improving existing evolutionary models (more on this topic later). Such studies support the view that MRI-based databases (fMRI, anatomy DTI, DSI) will provide a very important tool to study homologies between species in the future, allowing quantification (rather than just observation) of cross-species similarities and dissimilarities in brains. Neurophysiological experiments in the monkey (e.g. concerning connectivity between relevant brain regions, or concerning the functional role of putative mirror regions in mirror-type behaviors) will remain crucial to fuller understanding of the human mirror system and grounding processing models thereof. We see again the need for databasing homologies. Modelers (especially in linking skilled action and motor control to speech and language) must work across species – taking physiological hints from (usually) macaque studies to make sense of fMRI data from humans, to develop models of these processes in human brain areas.

Synthetic brain imaging (e.g., Arbib et al. 2000; Bonaiuto and Arbib 2014; Horwitz et al. 1999a) provides one approach

to using simulations of processing models of neural circuit to predict brain imaging results, PET and fMRI. An important technique in fMRI is structural equation modeling, SEM (McIntosh 2012), as a means to determine “effective connectivity.” To over-simplify, if fluctuations in activity in region A when executing a specific task are strongly correlated with those in region B, then if certain further conditions hold, we may infer that activity in A helps drive activity in region B *in that task*, and further processing may establish whether that “effective connection” is direct or follows a path through other regions. Such paths presumably demand the existence of anatomical connections; it is then a challenge for processing models to explain why an anatomical connection should be effective in one task but not in another. Future versions of synthetic brain imaging may be able to address this by building models grounded on anatomical connectivity and neurophysiological activity and using them to predict SEM results (Horwitz et al. 1999b). We may also note a number of important contributions to the linkage of brain mapping and cognitive processes by Karl Friston and his colleagues (Friston 2002; Friston et al. 2007, 1997; Friston and Dolan 2010).

To explore how apes may learn new manual gestures for communication, Gasser et al. (2013), discussing *ontogenetic ritualization of primate gesture as a case study in dyadic brain modeling*, use an assessment of ontogenetic ritualization to clarify the physical interactions and neural changes that seem necessary to support the transition from action to gesture more generally, charting the model development and integration involved in passing from separate models of various brain mechanisms to the analysis of *dyadic* interactions of two brains of two agents interacting with each other. As a byproduct of this, they chart the challenge to existing models of the *macaque* mirror system (Bonaiuto et al. 2007; Bonaiuto and Arbib 2010; Oztop and Arbib 2002) as they explore the processes that must underlie the ability of great apes to learn new gestures in interaction with each other (Arbib et al. 2014a). Such modeling studies may elucidate why it is that human brains, but not monkey brains, can support imitation.

## Language

Much of our current understanding of language is based on work in psycholinguistics (Traxler and Gernsbacher 2006) where data include error rates and timing of responses to linguistic tasks, but not neural correlates. However, the results of such studies can constrain processing models even though the constituent processes are not neurally localized. Further pertinent data come from studies of language acquisition in children, and studies of language learning more generally. Neurolinguistics can use findings from psycholinguistics, but further seeks data from aphasia (how brain lesions impair language performance) and other clinical populations as well

as from fMRI and other cortical mapping methods that localize regions where the BOLD or PET signal is greater for a given task than a control task – but this does not preclude a role for other regions in the task, nor does it clarify whether (and if so, why) the increased activation is required for the task. Understanding how these multiple areas interact in real time constitutes the major challenge facing the development of processing models for neurolinguistics. In view of the very important role of timing in language processing – linguistic input inherently unfolds over time, with each new incoming element serving to further constrain or elaborate the current representation – event-related brain potentials (ERPs) and other EEG- and MEG-based measures also play a particularly important role.

An important database for following the *acquisition* of language is given by CHILDES, the Child Language Data Exchange System (MacWhinney 2000), <http://childes.psy.cmu.edu/>, which is the child language component of the TalkBank system, <http://talkbank.org>. The goal of TalkBank is to foster fundamental research in the study of human and animal communication by constructing sample databases within each of the subfields studying communication. It will use these databases to advance the development of standards and tools for creating, sharing, searching, and commenting upon primary materials. Perhaps of most relevance to neuroinformatics is its AphasiaBank effort (MacWhinney et al. 2011), <http://talkbank.org/AphasiaBank/>, whose goal is construction of a shared database of multimedia interactions for the study of communication in aphasia – aiming, however, at the improvement of evidence-based therapy for aphasia rather than at processing models of language in which the effects of brain lesions can be simulated.

Modern psycholinguistics has reached a measure of consensus regarding many important features of language processing, and these features carry over into the (still very limited range of) neurobiologically constrained processing models (Bornkessel and Schlesewsky 2006; Friederici 2002; Hagoort 2005; Ullman 2004). In particular, most current models assign some role to competition between items (MacWhinney 1987; Vosse and Kempen 2000; Vosse and Kempen 2009) on the various linguistic levels (but see Clifton and Staub 2008 for explicit empirical counterarguments). All current models assume that processing occurs incrementally, trying to make role assignments and attachments as soon as possible (Kempen and Hoenkamp 1987) based on the relative strength of surface cues to interpretation (MacWhinney 2011). Because processing is competitive and cue-driven, it can maintain memory representations of incompletely assembled linguistic units for a certain period and it can tolerate a certain amount of sloppy interpretation (Ferreira et al. 2002). However, it has been proposed that storage of incompletely attached patterns entails processing costs (Gibson 1998; Hawkins 1999) and that this can lead to avoidance of certain structures. An alternative, more

recent, conceptualization of such “complexity” effects is that they derive from similarity-based interference during memory retrieval (e.g., Lewis et al. 2006; McElree et al. 2003).

These core features of language processing are modulated in interesting ways by cross-linguistic differences in language structure, as has been explored both psycho- and neurolinguistically (Bornkessel-Schlesewsky et al. 2011; Bornkessel-Schlesewsky and Schlesewsky 2009). Current models of the learning of lexical structure (Li et al. 2007; Silberman et al. 2007) derive from work with highly analytic languages such as English or Chinese. Such models may be unable to account for the intense interaction between lexicon, morphology, and syntax found in polysynthetic languages such as Inuit or Navajo. Even languages as closely related as English and German can show qualitative differences in the electrophysiological signatures for processes at the form-to-meaning interface (Bornkessel-Schlesewsky et al. 2011). The goal remains to formulate neurologically-grounded models of language processing. Some target areas for these models would include auditory and articulatory processing (Bohland et al. 2009b; Hickok 2009), lexical representation (Bohland et al. 2009b), procedural learning (Ullman 2004), and perspective-taking (Arbib 2012; MacWhinney 2008). However, the development of these models faces a number of serious empirical challenges. Data from the study of subjects with brain lesions have failed to isolate clear patterns of damage, partly because of the dynamic nature of language processing, and partly because of major differences in brain organization prior to stroke.

In addition to data from fMRI and ERP, cortico-cortical evoked potential studies (Matsumoto et al. 2005) can help narrow down the range of possibilities for such models. This special issue presents three models whose differing foci pose challenges for both model integration, and for understanding how to bring diverse data sets into a shared, integrated framework. Kempen (2013) models syntactic structure formation as an interactive activation and inhibition process running in a neural network and evaluates its success in accounting for selected psycholinguistic phenomena. Alday et al. (2013) build on their own fMRI and ERP data (Bornkessel-Schlesewsky and Schlesewsky 2009) from sentence comprehension tasks in a range of typologically diverse languages as part of a search for basic cognitive or neurobiological organizational principles to develop an initial computational model and show how it interacts with language-specific properties. Barrès and Lee (2013) offer an approach to language production rooted in construction grammar and extend it to comprehension. The approach is distinctive in that it relates utterances to abstract semantic representations (SemReps) of visual scenes rather than syntactic trees. Eye-tracking experiments provide a test of the model assumptions, while a dual path model in relation

to aphasia grounds their distinction between the heavy semantics of world knowledge and the light semantics of the categories of the grammar. The assumption, widely held by psycholinguists, that non-syntactic representations play a crucial role in language is shared by Alday et al. (2013) in their neurobiologically and cross-linguistically plausible model.

All this raises challenges for summarizing empirical fMRI and ERP data for modelers. In particular, we need a standard ontology for ERP recordings. At present, no tools are available to extend BrainMap-style meta-analyses to ERPs and this is clearly a disadvantage. The standard component labels such as “N400” and “P600” are problematic for modeling because there is a lot of leeway in their definition. Just using the N400 label gives no information about scalp distribution or the time course of the N400 component in a specific experimental condition. Thus, it seems necessary to provide somewhat more detailed information in order to allow database users to come up with their own categorization of an effect if desired. For this, information about timing and topography (as well as polarity) appears most crucial.

The Neural ElectroMagnetic Ontologies (NEMO) project (<http://nemo.nic.uoregon.edu/wiki/NEMO>) is a potential candidate. NEMO aims to create EEG and MEG ontologies and ontology based tools. MINEMO (Minimal Information for Neural ElectroMagnetic Ontologies, Frishkoff et al. 2011) extends MINI (Minimal Information for Neuroscience Investigations) to the ERP domain, with checklist terms explicated in NEMO. A related issue is to develop standards for extracting components from ERP signals and for linking them to a limited set of brain regions using, e.g., fMRI data.

However, language disabilities that may be under at least partial genetic control, such as stuttering (Fisher 2010) or Specific Language Impairment could provide a useful entry point for modeling (Fisher and Scharff 2009). New efforts are now enabling connections to be made between genes and gene expression and the higher-level organization of brain systems, addressing the challenges in unifying diverse human brain datasets. Bohland et al. (2013) include a discussion of how such data can be used to support and test models of brain mechanisms underlying language. A future challenge for neurolinguistics is to grapple with development (as opposed to acquisition and learning) as well, the key consideration being that the system is not constant across the time during which learning is taking place, but is rather also developing in many ways, partly under the control of genes and partly in response to the physical and social environment (Fisher and Ridley 2013).

### Challenges of Databasing Empirical Data

In developing neurobiologically realistic models of language, we need to integrate data from different methods. Specifically, given the rapidity of language processing, we cannot rely only

on fMRI but also need to consider methods with a high temporal resolution such as EEG and MEG. The integration between fMRI and ERP data thus appears particularly important, since these are the main two methods that have been used to glean data for neurolinguistic models. Source localization of ERPs in terms of dipoles does not seem to provide a good solution – even with a high number of electrodes, the inverse problem still persists. Here an inverse model is one that works back from a pattern of electrical activity detected at electrodes attached to the scalp to infer the pattern of cortical activity underlying it. Unfortunately, this problem is woefully ill-posed. Moreover, due to the different nature of the EEG and BOLD signals, constraining source localizations with neuroanatomical information from fMRI studies does not necessarily resolve the problem. Similar concerns hold for MEG (Sato et al. 2004) and estimating cortical current sources from EEG using near-infrared spectroscopy as a hierarchical prior has been suggested as a cheaper alternative (Aihara et al. 2012).

Thus, it appears that we need to look for alternative ways of integrating the different data types. In the spirit of this article, processing models could play a central role in linking spatial and temporal data. In addition to providing a functional link, these models may also allow us to seek paths from cortical models to ERPs, i.e., forward models in the sense of inferring the ERP from the cortical activity predicted by the model. Such efforts require processing models that use anatomy to constrain placement of neurons in 3D brain volumes, with estimates of propagation delays between different brain regions (Barrès et al. 2013).

Alternatively, it may be more promising to attempt to link EEG oscillations (i.e. variations of activity in different frequency bands) and BOLD activity. Indeed, several studies using concurrent EEG and fMRI measures have found that oscillatory activity can be used to predict BOLD activity on a trial-by-trial basis (Debener et al. 2006, 2005). From a different perspective, combining EEG with TMS may also help to provide spatial information for EEG data. Here, too, it may be helpful to attempt to work on specific problems within smaller focused communities before considering large-scale or universal connections.

In addition, there are many more general issues that have been addressed in the articles and editorial pages of this journal. Turning to neuroanatomical and related data to support processing models at the neurobiologically consistent neural network level, ideally we would want the information on how cells of type A projects to cells of type B. We also need timing information – not just region C goes to D, but how information flows from one region to another. This is currently possible (although not often done) for anatomically distinct cell populations, but this becomes far more challenging for functionally defined cell classes that might be spatially intermingled.

How can one find a format for *primary* data in a particular domain that experimentalists will agree upon? The reduced view of a dataset given in a published paper may be inadequate. One solution may be for journals to require entering data into domain-specific databases. Storing data as Supplementary Materials on a paper by paper basis is not nearly as useful since it does not require adoption of common ontologies which structure search and comparison. Such issues have long been a concern in neuroinformatics though usually focusing on specific data types, including some that we have introduced above. Another issue is that raw data storage may become uneconomical. The demands of storing and manipulating such data may be increasingly met, to the extent they can be met, by employing cloud architectures (Sobolev et al. 2012). In any case, modelers need access to the extended data set and the tools to extract “summary views” of the underlying data in a form amenable to the design and testing of processing models – and we have noted earlier (in the section [Bridging the Gap Between Models and Experiments](#)) the problems for modeling caused by differences in protocol and viewpoint in the empirical data.

While in an ideal world the data should be refereed, it is not practical – but refereeing a paper should suffice, allowing other scientists to check the details behind published claims as and when they see fit. In genetics the data are not reviewed in any serious way but there is a requirement to submit. However, solutions for genetics may not work for neuroscience since the field is too diverse – and the brain is much more difficult to describe than the mostly one-dimensional genome. Moreover, data description in a paper may entail accidental error. Kennedy (2012) argues that this can be avoided by preparing the raw data of a publication for data archive/data sharing at the time of publication even when there is no plan to share the data, but this assumes that a format for the given data type already exists *or is unambiguously defined in the process*, and this is too seldom the case in the neural study of action and language – and for much else in neuroscience. We want a system that allows people to locally define their own protocol in a very structured manner, but to globally be more diverse. SEDs in BODB can currently be just text (fMRI tables and connectivity data are the main exceptions) but a community using a shared workspace should define a data entry structure for each class of data they share.

Including raw data in a database enables users of the database to conduct their own analyses (i.e., apply other analysis methods if desired) and also has the advantage that it does not require one to “trust” other peoples’ choices in data analysis. On the other hand, raw data may contain unnecessary detail and raise the issue of reliable de-identification of individual datasets. In this second view, the level of detail included in published peer-reviewed datasets is enough. BrainMap provides an example of a database that works in this way and it attests to the utility of this type of

“high-level” data repository. It thus seems that summaries of published data provide a good starting point for data repositories that serve the purpose of informing both experimental work (meta-analyses as background information to inform experimental design and hypotheses) and modeling. However, under the assumption that the concerns regarding dataset de-identification can be resolved, it seems that, ultimately, at least some databases will need to provide access to raw (or very detailed) data.

It also seems important to find a viable tradeoff between standardizing the format of inputting/reporting data so as to make data extraction most efficient and not letting this format become too rigid, thus making it difficult to incorporate datasets that fall outside the mainstream. How does one link different summaries of the same set of experimental data? One approach might be to use hierarchical SEDs, providing for each class of empirical data a hierarchy of summary views of the data at increasing levels of granularity. The rationale for this approach is that different types of models need data at different levels of analysis. Making the hierarchy explicit would ensure that the fact that SEDs are linked to different views of the same dataset would be easily apparent, and makes sure that the link to original data is not lost. For example, a database dedicated to neurophysiology might store spike and event times tagged to anatomically localized cells and related to different task conditions. The lowest level summary might show a binned and summed histogram of event-triggered spike times with an inter-spike interval plot, the next level would smooth the histogram to generate views of neural firing rates related to an event., the next level up might show an average firing rate for the population, and the highest level might simply say whether firing increased or decreased from baseline in response to an event (maybe a normalized population firing rate). In this example, the models of macaque mirror neurons reviewed by Demiris et al. (2013) link to the highest level SED in the hierarchy, but it would be possible to inspect the hierarchy to generate targets for more detailed models.

### Challenges in Linking Models and Experiments

In this final section, we gather together a somewhat disparate set of observations, but all focused on the development of a neuroinformatics which involve the general concern that processing models must address diverse sets of data from the perspective offered by our concern with neural mechanisms supporting action, language, and the relations between them.

#### Testability of High-Level Computational Models

Many experimentalists seem to want models to be limited to those which make empirically tested or testable predictions.

However, there are two counter-arguments: (a) Models may be valuable for their implications for robotics applications (Demiris et al. 2013). (b) Predictions may also be evaluated for computational feasibility even when appropriate empirical methods for testing are not yet available. This does not settle the matter, but can determine which models remain viable as new experimental methods become available. In any case, modelers should understand enough to suggest which predictions might be ripe for testing (though experimentalists may come up with techniques that the modeler does not know about). This raises another challenge for databases of models: providing facilities to post predictions where experimenters will see them – fostering links to relevant data sets that exist, and encouraging experimenters to test predictions. This relates to the later discussion of “[Collaboratory Workspaces](#)” which can establish focused discussion with relevant modelers and/or other experimenters.

### How Should Models be Described for Databasing?

We want descriptions of models with function-driven components so that the modeler can break the model down into the essential elements that s/he wants to share with other modelers in a high-level representation. But what is the appropriate representation? The code for an actual implementation of a model (as in Model DB) should allow for very detailed model comparisons, but may obscure essential differences and similarities between the code at a higher level. Consider what happens as one moves from the NEURON environment to, e.g., neurolinguistic models. Model comparisons are difficult in general because of differences in scope and philosophy of the different models. Work will need to be done in order to find a format that permits such comparisons, with suggestions for guiding principles including the data that the model explains and its basic functional principles. One may expect that comparison tools will be rather domain-specific.

De Schutter (2008) addresses the question “Why are computational neuroscience and systems biology so separate?” and answers in part that in neuroscience many modelers still do not share their models freely, while this is standard practice in systems biology. At present, ModelDB offers a place where brain modelers can deposit their code along with instructions on how to use that code to conduct simulations that reproduce established results (and, of course, may also be used to generate new results with the model). By contrast, BODB seeks to catalog models in terms of summaries of the empirical data used to design and test them. As such, the two databases are complementary, and it is to be hoped that in future computational neuroscientists will increasingly provide linked descriptions of their models in both databases.

The INCF Program on Multiscale Modeling (<http://www.incf.org/programs/modeling>) offers links to a number of

initiatives that try to solve the model interoperability problem, but does not address the challenges of *cognitive neuroscience* that have been of central concern in this discussion paper. For example, the Network Interchange format for Neuroscience (NineML) is a markup language for model description designed to support specific implementations covering a wide range of modeling scales, but it focuses on networks of single compartment neurons. NeuroML (Gleeson et al. 2010; Goddard et al. 2001) offers an approach to providing a functional description of models which can be interpreted and imported into many different neural simulators. However, this currently lacks a high-level view of models such as that provided by BODB’s hierarchical module–submodule decomposition and can thus currently only capture neural models with a relatively large amount of biological detail (as captured in the NEURON and GENESIS models which dominate ModelDB).

Functional principles are also central to the issue of whether databases can help us to move towards neurobiologically plausible models of language. Note the earlier discussion of operating principles derived from psycholinguistics without reference to neural correlates. Most current models are essentially cognitive models which incorporate only a few, very high level, brain correlates. However, the “boxes” in these models typically map onto many regions, ignoring their complex connectivity. Furthermore, once we go beyond the input–output mapping, the relationship between the elements in the models and the elements in the brain are often unclear. It remains an open challenge to design databases that establish connections between these different levels, to support model development which maps functional principles onto neurobiologically realistic ones.

### Database Components

How do we capture what an experiment is about? Should we represent it with a free-form text description or within a structured framework? BrainMap has a taxonomy describing experiments (over 10,000 experiments from over 2,000 papers have been described with the taxonomy). The taxonomy has recently been developed further into CogPO (Fox et al. 2005; Turner and Laird 2012). But apart from labeling the experiments against a taxonomy, BrainMap uses “prose description” for the experimental conditions behind the stored brain scan data. The Brede Database description of experiments is close to that of BrainMap, labeling experiments in terms of the BrainMap taxonomy while also having free-form descriptions. On top of this, Brede Database labels each experiment against its own taxonomy, the so-called external components (Nielsen 2005). The taxonomies may evolve over time. Given a sufficient amount of data, data-driven analysis could possibly refine taxonomies.

Databases could foster post-publication review, and a rating facility would allow users to rate individual database items. Such functionalities seem not to be widespread among neuroinformatics databases, but the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC, <http://www.nitrc.org/>, Luo et al. 2009), a Web-based source for neuroimaging tools and resources (Kennedy 2010), supports reviews and ratings on “overall ease vs. functionality”, “download/installation ease” and “documentation/support quality”. For some of the pages describing scientific articles, the Brede Wiki maintains a section for critiques where post-publication reviews might be given. Reviews and ratings of publications, experimental results, or models could be used to provide a measure of confidence in those items. In sufficient numbers, these ratings might even aid in the scientific discovery process. For example, by using ratings as “weights” in meta-analyses, one might be able to leverage the expert knowledge of the scientific community in evaluating particular findings to devalue studies likely to have methodological or other problems and promote studies that appear to be the most sound. This process would formalize the existing process most researchers undertake manually in “weighting” the – often inconsistent – results found across the literature when surveying a particular neuroscientific domain. The downside (for both the machine and human versions) is that this may downplay the value of true innovations. In any case, the challenge of resolving inconsistencies in the literature remains. For example, two sets of data may appear inconsistent – but only because details of how the related protocols differed was not preserved in the databases.

Database updates and maintenance need to be considered: Does the data reflect current scientific knowledge? An example of a neuroinformatics database which has not been updated for a while is the fMRI Data Center and CoCoMac (<http://www.cocomac.org>), an online database for detailed information from tract-tracing connectivity studies in macaque monkeys (Kotter 2004). However, new developments on CoCoMac are under way (Bakker et al. 2011, <http://cocomac.g-node.org>). It may be worth noting that few neuroinformatics databases are (or will be?) complete, see, e.g. the case for coordinate-based databases, (Derrfuss and Mar 2009), though BrainMap has relatively good coverage (78 %) in terms of voxel-based morphometry studies (Laird et al. 2011).

### Federation

“Federation” refers to the challenging process of making different resources work together, either automatically and seamlessly, or through additional layers that integrate resources using one or more web service APIs. A new database may either copy and replicate the data from other

databases or call the original database through an application programming interface (API). If data is accessed through the online API, the new database loads up the original database server. This is not the case if the data are distributed. An important issue for us is that, for a specific processing model, the relevant SEDs may not be actual entries in a repository of primary data, but may each instead summarize a number of partial results extracted from such entries, structured in such a way as to best challenge modeling of specific brain functions. For example, BODB (Arbib et al. 2014b) takes this approach in federation with CoCoMac. CoCoMac describes data from studies that use injection of radiographic tracers that propagate in a retrograde (toward the soma) or anterograde (away from the soma) fashion along axons. Inspection of which regions are consequently labeled with the tracer allows inference of a region’s connectivity including an estimate of connectivity strength according to the intensity of the tracer at the labeled region. BODB provides connectivity SEDs that extract from this detailed information a more abstract representation – region X projects to region Y. The details of the projection such as the tracer type used and its estimated strength can be viewed via links from BODB’s connectivity SEDs to the corresponding CoCoMac entries. The summary representation in BODB can be used to create network graphs of selected regions which can then be applied to constrain a model’s architecture.

When federating databases, data need to be matched across systems. A common means for database federation is to provide deep linking, i.e., specific items in one database are linked to corresponding items in another. The widely used PubMed identifier is used across a large number of neuroinformatics databases and can thus be used to translate document identifiers between databases. However, it would be valuable to link each paper to more expansive resources. PubMed provides a LinkOut program (Marenco et al. 2008) to display links to external resources from their paper abstract view pages. Resource providers (such as neuroinformatics databases) provide them with a list of PubMed IDs and a way to generate URLs to their resources. This might provide an incentive for people to enter their data into neuroinformatics databases. For example, if one goes to the page for the MNS2 paper (Bonaiuto et al. 2007) on PubMed, <http://www.ncbi.nlm.nih.gov/pubmed/17028884>, and clicks on the Brain Operation Database on the “LinkOut – more resources” menu, one is transferred to the BODB page for that paper, <http://bodb.usc.edu/bodb/literature/pubmed/17028884/>, which in turn has links to related models (including the MNS2 model itself) and related simulation results. Other identifiers to translate between databases are spatial coordinate descriptions available in, e.g., BrainMap, Brede and now also in BODB. Conforming to the standards of the Semantic Web with RDF and “Linked Data” may also make it easier to integrate items across databases. A recent example of using Semantic Web

technologies is the Cognitive Paradigm Ontology (CogPO) based on the BrainMap taxonomy (Turner and Laird 2012). The Semantic Web Applications in Neuromedicine (SWAN) effort provides an ontology developed in the context of building a series of applications for biomedical researchers, as well as in extensive discussions and collaborations with the larger bio-ontologies community (Ciccarese et al. 2008).

Web services can provide views into a database as XML or JSON text for another web application or external program. This is a simple, alternative means of sharing across resources that does not require explicit connection to the underlying database or substantial knowledge of the underlying schema. Nonetheless, linking data can be problematic when different databases may employ different ontologies. If we use the same terms to mean different things, or different terms to mean similar things, we cannot easily answer questions that span across multiple resources. NeuroLex (<http://neurolex.org/wiki/>) is a dynamic lexicon of 22,292 neuroscience terms, including 289 neurons and 940 brain parts. The aim is to remedy the inconsistent use of terminology in databases and the literature. However, this work on constructing ontologies to help organize neuroscience concepts into category hierarchies, etc., appears to be at an early stage. Bandrowski et al. (2012) provide a preliminary report on exploring mammalian brain connectivity using NeuroLex.

Brain regions as database items may be linked across databases, e.g., the BrainNavigator and Brede Wiki link to the brain regions of BAMS. Future federation may be possible between BAMS and brain regions defined by spatial profiles of gene expression in the mouse or other species (see Bohland et al. 2013), as well as between BODB and BAMS. The methods set forth in Knowledge Engineering from Experimental Design, KefED (Russ et al. 2011), may be used to curate for BAMS and possibly also BODB. As the BrainMap schema (Fox et al. 2005) is developed into CogPO it may be more relevant for use in other tools, e.g., KefED. BODB has made some initial efforts to include Brede Database data about papers, experiments and coordinates by copying the Brede Database from its XML distribution. The present version of BODB links back with deep links from the “Summary of Experimental Data” to the experiments in the Brede Database. The Brede Wiki in turn links back to BODB’s “Summary of Experimental Data” and “Literature” (and Brede Database) through hard-coded deep links.

The existence of diverse neuroinformatics databases opens up numerous possibilities to link and query between them. Linking between items in the databases may not be straightforward as they can have “fuzzy” relationships, e.g., as shown in the Online Brain Atlas Reconciliation Tool (OBART) (Bohland et al. 2009a) where corresponding brain regions in different atlases often do not overlap completely. OBART enables users to explore relationships between multiple existing MRI-based human brain atlases; one may

recall Bota’s work reconciling macaque atlases as part of his NHDB (Bota and Arbib 2004). (Intriguingly, the Website <http://obart.info> is devoted to *Le buzz du poker gratuit*. However, the current version of OBART can be found at both <http://obart.brainarchitecture.org> and <http://qnl.bu.edu/obart>.)

With the explosion of Web-based resources in the neurosciences, *mining* these resources is increasingly challenging without suitable application programming interfaces (APIs) and systems to support machine to machine interactions. The need for so-called *web services* has been recognized in the bioinformatics community (Stein 2002), where data in large distributed repositories are routinely made machine accessible. In neuroscience, database interoperability is currently lagging, with many resources containing isolated data that can only be accessed through manual interaction with a web browser. The Neuroscience Information Framework (NIF, Gardner et al. 2008, [www.neuinfo.org/](http://www.neuinfo.org/)) has made some progress along these lines by providing the DISCO framework (Marenco et al. 2010) to allow individual databases to make particular resources available to the NIF search engine. NIF allows one to search for items over a host of neuroinformatics databases, but seems ill-structured – a search engine overlay for a range of resources, rather than a federation. As noted by Arbib et al. (2014b): “A search for ‘cerebellum model. yielded 1,214,963 results, most of them irrelevant to computational modeling ... [but] the search yielded 17 distinct results when filtered for the ‘Models’ data type. These models are sourced only from ModelDB and Open Source Brain.” Immediate improvements could include a capacity for more finely structured search and tools to perform inference across resources. More generally, there is a need for database developers to make data available through simple, standardized APIs (or direct exposure of database connections) to eliminate the common practice of downloading and maintaining complete local copies of remote resources through “screen scraping” or direct transfer. Such services can be implemented, for example, using REST (Representational state transfer) compliant web services that expose data in text-based formats such as XML or JSON, and for which there exist implementations in all modern web application frameworks.

Elsevier’s BrainNavigator ([www.brainnav.com/](http://www.brainnav.com/)) is an online product containing content and toolsets built around stereotaxic brain atlases across multiple species. It contains parcellations and annotations of up to 1,200 unique structures in each species, and drawings with histology to match, together with tools to visualize in 3D and to overlay against researcher histology to aid in neuroanatomical interpretation. The nomenclature used is that of Luis Puelles, Charles Watson, and George Paxinos, and the core database of structures also has APIs for bi-directional linking based on either structure or stereotaxic coordinate. Any partner site with

structure-based information (such as BAMS, BODB, NeuroLex) can be linked to or from BrainNavigator on a per-structure basis, and every structure for each species has its own open URL (to an atlas thumbnail summary) for direct linking in. The coordinate-based API (REST XML) is in pilot stage, providing where-am-I? lookup, returning structure name and hierarchy from coordinate input.

### Collaboratory Workspaces

Finally, we turn to the issue of encouraging collaboration between modelers and experimentalists who share a passion for the same area of neuroscience, though from different perspectives. As emphasized in several editorials in *Neuroinformatics* (Ascoli 2006; Kennedy 2006; Kennedy et al. 2011), experimenters need to have an incentive to share their data – and modelers their models – in order for databases to be potentially useful tools. Again, BrainMap provides a good example for how this could work: by entering their data into the database, users are afforded the possibility of conducting meta-analyses relevant to their own research. In general, then, the tools provided in a database appear to provide a good potential source of motivation for researchers to share their data. Furthermore, the inclusion of data in meta-analyses leads to citations of the original publication, which is another source of motivation for data-sharing. There may also be additional incentives, and specifically ones that might be applicable beyond the meta-analysis scenario type described above. For repositories including raw data sharing, one option may be to track the number of times a dataset is downloaded as an index.

An alternative option to encouraging people to share their data may be to harvest data summaries from journal publications. To this end, some of the leading journals in the field (*Human Brain Mapping*, *Neuroimage*, possibly others) are currently devising a standardized input matrix for entering data, which will make it possible to extract this information in a relatively automatic fashion. However, this of course only allows for the harvesting of relatively abstract data summaries and not of raw data. Web-based technologies have long promised to change the face of scientific collaboration (Kouzes et al. 1996; Sagotsky et al. 2008), with so-called Web 2.0 concepts – including increased information sharing, blogging, and social networking – offering new means for increased scientific productivity while raising other concerns about quality and intellectual property (Waldrop 2008). A critical challenge for new neuroinformatics efforts is to facilitate meaningful collaborations, particularly between modelers and experimentalists. Social web technologies, mimicking the interactions researchers have at conferences and workshops through the use of wikis, blogs, forums, and real-time chat tools may help to facilitate adoption of such resources.

The issue of data and database ownership is a non-technical issue which may hinder integration between databases.

Maintaining ownership of experimental data or of the databases which make them accessible may be important for acquiring grants to fund the continued research. Exposing the data through an API or distributing an entire database may dilute the possibility, e.g., of co-authorship used to prove the utility of the data or database for funding bodies. A potential strength of databases may lie in enabling smaller groups to engage in joint projects sharing data and modeling ideas among themselves rather than with the general neuroscience public. This type of collaboration, for which BODB aims to offer a good platform (see the discussion of *Collaboratory Workspaces* in Arbib et al. 2014b), may be particularly useful in encouraging data-model integration on particular topics of interest and could provide a first step towards larger collaborative efforts. It also appears plausible that researchers may be more highly motivated to share their data (including raw data) for the purposes of a concrete, smaller-scale project of this type as opposed to the more abstract goal of increasing the possibility of additional citations by inputting data into a database.

### Information Sharing Statement

See [Appendix](#) for shared resources related to this work.

### Acknowledgement, Discussion Groups and Workshop Participants

The present paper is based on discussions held on Wednesday July 27, as part of the Workshop on “Action, Language and Neuroinformatics” held on July 25–27, 2011, in Los Angeles under the aegis of the USC Brain Project of the University of Southern California, and organized by Michael A. Arbib. The Workshop was supported in part by the National Science Foundation under Grant No. 0924674.

There were six discussion groups, three (1a,b,c) meeting in parallel in the morning and three (2a,b,c) in the afternoon. The participants of the Workshop were divided into three groups, one each on Action, Language, and Neuroinformatics. The discussion groups were then organized as follows. The name of the rapporteur for each session is marked in **Bold**. The present paper is based on the integration of their six reports.

*1a. Half the Action group + half the Neuroinformatics group:* assessing the model-data integration and neuroinformatics needs of the Action group: **Bonaiuto**, Oztop, Demiris, Vanduffel, Arbib, Marques, Bohland.

*1b. Half the Neurolinguistics group + the other half of the Neuroinformatics group:* assessing the model-data integration and neuroinformatics needs of the Neurolinguistics group: **Bornkessel-Schlesewsky**, Small, MacWhinney, Miikkulainen, Nielsen, Fox, Barrès, Schuler.

*1c. The other half of the Action group + the other half of the Neurolinguistics group:* Defining shared modeling challenges and the development of a shared conceptual framework. **Kemmerer**, Aziz-Zadeh, Cartmill, Gasser, Grosvald, Wood, Kempen, Lee, Schilling.

*2a. Action group:* What are the key data and/or conceptual issues ripe for modeling; what are the key lines of modeling that hold most promise to address these data/issues? **Oztop**, Demiris, Vanduffel, Cartmill, Arbib, Aziz-Zadeh, Gasser, Schilling, Wood.

*2b. Neurolinguistics group:* What are the key data and/or conceptual issues ripe for modeling; what are the key lines of modeling that hold most promise to address these data/issues? **MacWhinney**, Kempen, Grosvald, Small, Bornkessel-Schlesewsky, Miikkulainen, Kemmerer, Lee, Barres

2c. *Neuroinformatics group*: What tools are ripe for sharing, or should be ripened? What are promising lines for federation? **Nielsen**, Marques, Bohland, Bonaiuto, Bota, Fox, Schuler.

Brief biosketches of the participants, access to selected papers, and abstracts of their talks may be found at the Workshop Website: [http://uscbp.usc.edu/mediawiki/index.php/2011\\_Workshop](http://uscbp.usc.edu/mediawiki/index.php/2011_Workshop).

## Appendix: A Tabulation of Present and Future Resources

**Table 1** Neuroinformatics resources

Name and URL	Scope	Notes
Allen Brain Atlas <a href="http://www.brain-map.org">www.brain-map.org</a>	Gene expression data from the mouse brain, human brain, developing mouse brain, mouse spinal cord, and aggregations of other gene expression data. Gene expression patterns in the mouse brain	There are adult and developing human atlases, a developing mouse atlas, and a macaque atlas here as well.
Aphasia Bank <a href="http://talkbank.org/AphasiaBank/">http://talkbank.org/AphasiaBank/</a>	Multimedia interactions for the study of communication in aphasia	This database is focused on use by clinicians.
Brain Architecture Management System, BAMS <a href="http://brancusi1.usc.edu/">http://brancusi1.usc.edu/</a>	Information about neural circuitry	Supplies inference engines and five interrelated modules: Brain Parts (gray matter regions, major fiber tracts, and ventricles), Cell Types, Molecules, Connections (between regions and cell types), and Relations (between parts identified in different neuroanatomical atlases).
Brain Operation Database, BODB <a href="http://nsl.usc.edu/bodb/">http://nsl.usc.edu/bodb/</a>	Links processing models with summaries of empirical data	In addition to comparison of empirical data with simulation results, BODB accesses models via both a structural ontology (brain structures) and a functional ontology (brain operating principles, BOPs) and offers tools for model comparison
BrainMap <a href="http://www.brainmap.org/">http://www.brainmap.org/</a>	Repository for human fMRI brain imaging data.	Brain imaging data is linked to a rich ontology for the description of experimental procedures. Provides statistical tools (activation likelihood estimation, ALE) for meta-analysis of coordinate based activation data.
Brede Wiki <a href="http://neuro.imm.dtu.dk/wiki/">http://neuro.imm.dtu.dk/wiki/</a>	Contains information from published scientific articles and ontologies for describing that information Contains primarily information from published peer-reviewed neuroimaging articles	Data and text can be entered with raw Wiki text entry (as in Wikipedia). Web services allow for meta-analysis and coordinate-based information retrieval.
CoCoMac, Collations of Connectivity data on the Macaque brain <a href="http://www.cococmac.org">http://www.cococmac.org</a>	Systematic record of the known wiring of the primate brain. The main database contains details of hundreds of tracing studies in their original descriptions.	BODB federates with CoCoMac to provide summarized accounts of cortical connectivity with links back to the original CoCoMac entries for further information if desired.
ConnectomeDB, the Human Connectome Project <a href="http://www.humanconnectome.org/">http://www.humanconnectome.org/</a>	NIH funded large scale project to generate a database of the human brain connectome. It includes both anatomical white matter tracts defined by Diffusion Tensor Imaging and functional connectivities based on resting state fMRI analysis.	
Conte69 Atlas <a href="http://brainvis.wustl.edu/wiki/index.php/Caret:Atlases:Conte69_Atlas">http://brainvis.wustl.edu/wiki/index.php/Caret:Atlases:Conte69_Atlas</a>	Human surface-based brain atlas	A tool to quantitatively define anatomical brain regions.
MEG-SIM <a href="http://cobre.mrn.org/megsim/">http://cobre.mrn.org/megsim/</a>	Shared database for simulated and recorded MEG data.	Still at an early stage. The stated goal is to create “realistic simulated data sets in formats used by each of the 3 major MEG manufacturers. These can then be directly tested using various algorithms which include multidipole, spatiotemporal modeling, current reconstruction, minimum norm, and beamforming methods.”

Neuroinform

**Table 1** (continued)

Name and URL	Scope	Notes
ModelDB <a href="http://senselab.med.yale.edu/modeldb/">http://senselab.med.yale.edu/modeldb/</a>	Documented code for models, primarily those implemented in the NEURON and GENESIS modeling environments	For each model, provides instructions on how to run the code to get published results.
Neural ElectroMagnetic Ontologies, NEMO <a href="http://nemo.nic.uoregon.edu/wiki/NEMO">http://nemo.nic.uoregon.edu/wiki/NEMO</a>	EEG and MEG ontology for ERP databasing.	Aims at providing statistical tools for analysis of EEG and MEG activations patterns. So far, few entries are available.
Neuroscience Information Framework, NIF <a href="http://www.neuinfo.org/">www.neuinfo.org/</a>	Allows one to search for items over a host of neuroinformatics databases.	Seems ill-structured – a search engine overlay for a range of resources, rather than a federation
Neurosynth <a href="http://neurosynth.org">http://neurosynth.org</a>	Automatically synthesizes results of neuroimaging studies	Extracts activation coordinates and high-frequency terms from published articles and generates statistical maps based on keyword co-occurrence with coordinate ranges.
Online Brain Atlas Reconciliation Tool, OBART <a href="http://obart.brainarchitecture.org">http://obart.brainarchitecture.org</a> <a href="http://qnl.bu.edu/obart">http://qnl.bu.edu/obart</a>	Human brain regions defined by multiple MRI-based atlases	Allows comparison of brain regions based on spatial definitions rather than terms or labels.
Surface Management System Database, SumsDB <a href="http://sumsdb.wustl.edu/sums/index.jsp">http://sumsdb.wustl.edu/sums/index.jsp</a>	Repository of brain-mapping data (surfaces & volumes; structural & functional data)	It accepts unpublished data (which however is not public). Provides WebCaret tools for visualization.

**Table 2** Other resources

Name and URL	Scope	Notes
CHILDES, the Child Language Data Exchange System, <a href="http://childes.psy.cmu.edu/">http://childes.psy.cmu.edu/</a>	Resources for following the <i>acquisition</i> of language	CHILDES is the child language component of the TalkBank system.
Cognitive Paradigm Ontology, CogPO <a href="http://cogpo.org/">http://cogpo.org/</a>	Ontology for the description of experimental paradigms in cognitive science	
FrameNet (Fillmore 1976) <a href="https://framenet.icsi.berkeley.edu/fndrupal/">https://framenet.icsi.berkeley.edu/fndrupal/</a>	Natural Language Processing (NLP) oriented frame based conceptual ontology.	Other NLP resources: WordNet (Miller 1995) <a href="http://wordnet.princeton.edu/">http://wordnet.princeton.edu/</a> VerbNet ( <a href="http://verbs.colorado.edu/~mpalmer/projects/verbnet.html">http://verbs.colorado.edu/~mpalmer/projects/verbnet.html</a> ) PropBank (Palmer 2009) ( <a href="http://verbs.colorado.edu/~mpalmer/projects/ace.html">http://verbs.colorado.edu/~mpalmer/projects/ace.html</a> ), The goal of (Palmer 2009) is to link all these resources. The more general goal of linking language resources is also a major part of the Linking Open Data Cloud. See: <a href="http://richard.cyganiak.de/2007/10/lod/">http://richard.cyganiak.de/2007/10/lod/</a>
Purdue ASL Database <a href="http://www2.ece.ohio-state.edu/~aleix/ASLdatabase.htm">http://www2.ece.ohio-state.edu/~aleix/ASLdatabase.htm</a>	Database of American Sign Language (ASL) videos and transcriptions	Still very preliminary but could be scaled up.

**Table 3** A target set of new neuroinformatics resources

Scope	Notes
Brain development data	Such data are hard to find, and can play a role as models start to incorporate genetics data and take seriously the issues of the “wetware” changing during long-term learning/continued performance of complex behaviors
Conceptual model ontology and database	To this day, no database exists that offers an ontology to store and compare conceptual models. BODB and ModelDB are both oriented towards computational models and the challenge would be more of integration rather than having one of these databases take over the role to store conceptual models.
EEG and MEG visualization tools and standardized head models for display.	Similar to the Talairach Daemon visualization tool ( <a href="http://www.talairach.org/daemon.html">http://www.talairach.org/daemon.html</a> ) for tomographic imaging, a standard head for display of EEG and MEG results would benefit the neurolinguistic community both in terms of the ease of visual comparison and in providing a way to report EEG and MEG results in papers in a more uniform table-like way as it has become standard for fMRI results (using Talairach or MNI coordinates).
Electrocorticography (ECoG) data	These data, gleaned from electrodes placed directly on the exposed surface of the brain to record electrical activity from the cerebral cortex, are becoming increasingly important in speech/language models
Event-related potential (ERP) data linked to linguistic and other cognitive tasks.	This requires a standard ontology for ERP recordings. One candidate is provided by the Neural ElectroMagnetic Ontologies (NEMO) project, <a href="http://nemo.nic.uoregon.edu/wiki/NEMO">http://nemo.nic.uoregon.edu/wiki/NEMO</a> , that aims to create EEG and MEG ontologies and ontology based tools. A related issue is to develop standards for linking extracting components from ERP signals and linking them to a limited set of brain regions using, e.g., fMRI data. (See above note on EEG and MEG visualization tools.)
Gestures used both by primates and by humans	This requires annotated videotapes showing the context and dyadic interactions in which the gestures occur. For humans, data supporting analysis of linguistic/paralinguistic (sign and cospeech gesture) vs non-linguistic hand motions would be useful. Motion Capture would yield even finer data than videotaping but is more limited in its applicability and would require extensive processing (e.g., shape and motion recognition) to yield data that might then be a basis for actual interpretation. Longitudinal data would be especially useful.
Lesion database linked to MRI and experimental paradigm database.	The study of linguistic performances of brain damaged patients and the development of brain based models from this data would benefit from the creation of a lesion database similar in format to BrainMap. Crucially, such a database should not be limited to aphasics or to other specific impairments since brain damage analysis would benefit from a better understanding of the loss of functions as whole rather than multiplying the smaller scopes analyses. An example of this would be the comparative analysis of aphasia and apraxia in a database to better understand the link between language and action systems. This would gather data relevant to modeling neural mechanisms of action and language, not being limited to the clinical emphasis of AphasiaBank
Neurohomology data – NHDB revisited	A resource (replacing the no longer active NHDB) to integrate data on human brains and the brains of other species to establish homologies that support new hypotheses about detailed neurophysiological, neurochemical and genetic mechanisms that are hard to resolve from the human data alone. MRI-based databases (fMRI, anatomy (T1, T2), DTI, DSI,) will provide an important tool to study homologies between species in the future, allowing one to quantify (rather than just observe) cross-species similarities and dissimilarities in brains. These can then be integrated with gene expression data and with macaque (and other animal) neurophysiology.
Neurophysiological data linked to behavior	Need the data organized so that one may extract summaries showing parametric variation across similar tasks under similar conditions. The German Neuroinformatics Node offers some tools and services (see <a href="http://www.g-node.org/tools-and-services">http://www.g-node.org/tools-and-services</a> ) that may be relevant, including a standard format for data annotation in electrophysiology, and a common terminology (based on the odML metadata approach) to facilitate data sharing and reproducibility of experiments. Ideally the neural responses and the task would have a hierarchy of descriptions at increasing levels of granularity. Sharing neurophysiological data would also require increased used of standardized coordinate frames.
Neurophysiological data linked to dyadic interactions	For <i>social</i> neuroscience to advance, ways must be found to track neural activity in conspecifics as they interact with one another.

## References

- Aihara, T., Takeda, Y., Takeda, K., Yasuda, W., Sato, T., et al. (2012). Cortical current source estimation from electroencephalography in combination with near-infrared spectroscopy as a hierarchical prior. *NeuroImage*, *59*, 4006–4021.
- Alday, P., Schlesewsky, M., Bornkessel-Schlesewsky, I. (2013). Towards a computational model of actor-based language comprehension. *Neuroinformatics*. doi:10.1007/s12021-013-9198-x.
- Arbib, M. A. (1981). Perceptual structures and distributed motor control. In V. B. Brooks (Ed.), *Handbook of physiology — the nervous system II. Motor control* (pp. 1449–1480). Bethesda: American Physiological Society.
- Arbib, M. A. (2012). *How the brain got language: the mirror system hypothesis*. New York & Oxford: Oxford University Press.
- Arbib, M. A., & Bota, M. (2003). Language evolution: neural homologies and neuroinformatics. *Neural Networks*, *16*, 1237–1260.
- Arbib, M. A., Billard, A. G., Iacoboni, M., & Oztop, E. (2000). Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Networks*, *13*, 975–997.
- Arbib, M. A., Ghanesh, V., Gasser, B. (2014a). Dyadic brain modeling, ontogenetic ritualization of gesture in apes, and the contributions of primate mirror neuron systems. *Philosophical Transactions of the Royal Society B*, to appear.
- Arbib, M. A., Plangprasopchok, A., Bonaiuto, J. J., Schuler R. E. (2014b). A Neuroinformatics of brain modeling and its implementation in the Brain Operation Database BODB. *Neuroinformatics*. doi:10.1007/s12021-013-9209-y.
- Ascoli, G. A. (2006). The ups and downs of neuroscience shares. *Neuroinformatics*, *4*, 213–216. doi:10.1385/NI:4:3:213.
- Aziz-Zadeh, L., Sheng, T., Liew, S.-L., Damasio, H. (2011). Understanding otherness: the neural bases of action comprehension and pain empathy in a congenital amputee. *Cerebral Cortex*, *4*, 811–819.
- Bakker, R., Potjans, T. C., & Diesmann, M. (2011). Macaque structural connectivity revisited: CoCoMac 2.0. *BMC Neuroscience*, *12*(supplement 1), 72.
- Bandrowski, A., Gupta, A., Gupta, A., Larson, S., Martone, M. (2012). Exploring mammalian brain connectivity using NeuroLex. *Neuroinformatics 2012 Abstracts*: <http://www.neuroinformatics2012.org/abstracts/exploring-mammalian-brain-connectivity-using-neurolex>.
- Barrès, V., Lee, J. (2013). Template construction grammar: from visual scene description to language comprehension and agrammatism. *Neuroinformatics*. doi:10.1007/s12021-013-9197-y.
- Barrès, V., Simons, A., & Arbib, M. A. (2013). Synthetic event-related potentials: a computational bridge between neurolinguistic models and experiments. *Neural Networks*, *37*, 66–92.
- Bohland, J. W., Bokil, H., Allen, C. B., & Mitra, P. P. (2009a). The brain atlas concordance problem: quantitative comparison of anatomical parcellations. *PloS One*, *4*, e7200.
- Bohland, J. W., Bullock, D., & Guenther, F. H. (2009b). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience*, *22*, 1504–1529.
- Bohland, J., Myers, E., Kim, E. (2013). An informatics approach to integrating genetic and neurological data in speech and language neuroscience. *Neuroinformatics*. doi:10.1007/s12021-013-9201-6.
- Bonaiuto, J. J., & Arbib, M. A. (2010). Extending the mirror neuron system model, II: what did I just do? A new role for mirror neurons. *Biological Cybernetics*, *102*, 341–359.
- Bonaiuto, J. J., & Arbib, M. A. (2014). Modeling the BOLD correlates of competitive neural dynamics. *Neural Networks*, *49*, 1–10.
- Bonaiuto, J. B., Rosta, E., & Arbib, M. A. (2007). Extending the mirror neuron system model, I : audible actions and invisible grasps. *Biological Cybernetics*, *96*, 9–38.
- Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*, 787–821.
- Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: a cross-linguistic approach. *Language and Linguistics Compass*, *3*, 19–58.
- Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., et al. (2011). Think globally: cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language*, *117*, 133–152.
- Bota, M., & Arbib, M. A. (2004). Integrating databases and expert systems for the analysis of brain structures: connections, similarities, and homologies. *Neuroinformatics*, *2*, 19–58.
- Bota, M., & Swanson, L. W. (2008). BAMS neuroanatomical ontology: design and implementation. *Front Neuroinformatics*, *2*, 2. doi:10.3389/neuro.11.002.2008.
- Ciccarese, P., Wu, E., Wong, G., Ocana, M., Kinoshita, J., et al. (2008). The SWAN biomedical discourse ontology. *Journal of Biomedical Informatics*, *41*, 739–751.
- Clifton, C., & Staub, A. (2008). Parallelism and competition in syntactic ambiguity resolution. *Language and Linguistics Compass*, *2*, 234–250.
- Craik, K. J. W. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.
- De Schutter, E. (2008). Why Are computational neuroscience and systems biology so separate? *PLoS Computational Biology*, *4*, e1000078.
- Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience*, *25*, 11730–11737.
- Debener, S., Ullsperger, M., Siegel, M., & Engel, A. K. (2006). Single-trial EEG-fMRI reveals the dynamics of cognitive function. *Trends in Cognitive Sciences*, *10*, 558–563.
- Demiris, Y., Aziz-Zadeh, L. A., Bonaiuto, J. J. (2013). Information processing in the mirror neuron system in primates and machines. *Neuroinformatics*. doi:10.1007/s12021-013-9200-7.
- Denys, K., Vanduffel, W., Fize, D., Nelissen, K., Peuskens, H., et al. (2004). The processing of visual shape in the cerebral cortex of human and nonhuman primates: A functional magnetic resonance imaging study. *Journal of Neuroscience*, *24*, 2551–2565.
- Derrfuss, J., & Mar, R. A. (2009). Lost in localization: the need for a universal coordinate database. *NeuroImage*, *48*, 1–7.
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*, 176–180.
- Ferreira, F., Ferraro, V., & Bailey, K. G. D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11–15.
- Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, *280*, 20–32.
- Fisher, S. (2010). Genetic susceptibility to stuttering. *New England Journal of Medicine*, *362*, 750–752.
- Fisher, S. E., & Ridley, M. (2013). Culture, genes, and the human revolution. *Science*, *340*, 929–930.
- Fisher, S., & Scharff, C. (2009). FOXP2 as a molecular window into speech and language. *Trends in Genetics*, *25*.
- Fox, P. T., Laird, A. R., Fox, S. P., Fox, P. M., Uecker, A. M., et al. (2005). BrainMap taxonomy of experimental design: description and evaluation. *Human Brain Mapping*, *25*, 185–198.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*, 78–84.
- Frishkoff, G. A., Sydes, J., Mueller, K., Frank, R., Curran, T, et al. (2011). Minimal Information for Neural Electromagnetic Ontologies

- (MINEMO): a standards-compliant method for analysis and integration of event-related potentials (ERP) data. *Standards in Genomic Sciences*, 5(2): 211–223.
- Friston, K. J. (2002). Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annual Review of Neuroscience*, 25, 221–250.
- Friston, K. J., & Dolan, R. J. (2010). Computational and dynamic models in neuroimaging. *NeuroImage*, 52, 752–765.
- Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, 6, 218–229.
- Friston, K. J., Ashburner, J., Kiebel, S. J., Nichols, T. E., Penny, W. D. (Eds.) (2007). *Statistical parametric mapping: the analysis of functional brain images*. London: Academic Press.
- Fujii, N., Hihara, S., & Iriki, A. (2008). Social cognition in premotor and parietal cortex. *Social Neuroscience*, 3, 250–260.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119, 593–609.
- Gardner, D., Akil, H., Ascoli, G., Bowden, D., Bug, W., et al. (2008). The neuroscience information framework: a data and knowledge environment for neuroscience. *Neuroinformatics*, 6, 149–160.
- Gasser, B., Cartmill, E., & Arbib, M. A. (2013). Ontogenetic ritualization of primate gesture as a case study in dyadic brain modeling. *Neuroinformatics*. doi:10.1007/s12021-013-9182-5.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Gleeson, P., Crook, S., Cannon, R. C., Hines, M. L., Billings, G. O., et al. (2010). NeuroML: a language for describing data driven models of neurons and networks with a high degree of biological detail. *PLoS Computational Biology*, 6, e1000815.
- Goddard, N. H., MI, H., Howell, F., Cornelis, H., Shankar, K., & Beeman, D. (2001). Towards NeuroML: model description methods for collaborative modelling in neuroscience. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356, 1209–1228.
- Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, 9, 416–423.
- Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language & Cognitive Processes*, 75, 244–285.
- Hickok, G. (2009). The functional neuroanatomy of language. *Physics of Life Reviews*, 6, 121–143.
- Horwitz, B., Tagamets, M. A., & McIntosh, A. R. (1999a). Neural modeling, functional brain imaging, and cognition. *Trends in Cognitive Sciences*, 3, 91–98.
- Horwitz, B., Tagamets, M. A., & McIntosh, A. R. (1999b). Neural modeling, functional brain imaging, and cognition. *TRENDS in Cognitive Sciences*, 3, 91–98.
- Jones, A. R., Overly, C. C., & Sunkin, S. M. (2009). The allen brain atlas: 5 years and beyond. *Nature Reviews. Neuroscience*, 10, 821–828.
- Kempen, G. (2013). Prolegomena to a neurocomputational architecture for human grammatical encoding and decoding. *Neuroinformatics*. doi:10.1007/s12021-013-9191-4.
- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11, 201–258.
- Kennedy, D. N. (2006). Where's the beef? Missing data in the information age. *Neuroinformatics*, 4, 271–273. doi:10.1385/NI:4:4:271.
- Kennedy, D. N. (2010). Making connections in the connectome era. *Neuroinformatics*, 8, 61–62.
- Kennedy, D. N. (2012). The benefits of preparing data for sharing even when you don't. *Neuroinformatics*, 10, 223–224.
- Kennedy, D. N., Ascoli, G. A., & De Schutter, E. (2011). Next steps in data publishing. *Neuroinformatics*, 9, 317–320. doi:10.1007/s12021-011-9131-0.
- Kotter, R. (2004). Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. *Neuroinformatics*, 2, 127–144.
- Kouzes, R. T., Myers, J. D., & Wulf, W. A. (1996). Collaboratories: doing science on the Internet. *Computer*, 29, 40–46.
- Laird, A. R., Lancaster, J., & Fox, P. T. (2005). BrainMap: the social evolution of a human brain mapping database. *Neuroinformatics*, 3, 65–77.
- Laird, A. R., Eickhoff, S. B., Kurth, F., Fox, P. M., Uecker, A. M., et al. (2009). ALE meta-analysis workflows via the brainmap database: progress towards a probabilistic functional brain atlas. *Front Neuroinformatics*, 3, 23. doi:10.3389/neuro.11.023.2009.
- Laird, A. R., Eickhoff, S. B., Fox, P. M., Uecker, A. M., Ray, K. L., et al. (2011). The BrainMap strategy for standardization, sharing, and meta-analysis of neuroimaging data. *BMC Research Notes*, 4, 349.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10, 447–454.
- Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, 31, 581–612.
- Lohrey, J. M., Killeen, N. E. B., & Egan, G. F. (2009). An integrated object model and method framework for subject-centric e-Research applications. *Frontiers in Neuroinformatics*, 3, 19.
- Luo, X.-Z., Kennedy, D., & Cohen, Z. (2009). Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC) resource announcement. *Neuroinformatics*, 7, 55–56.
- MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale: Lawrence Erlbaum.
- MacWhinney, B. (2000). *The chldes project: tools for analyzing talk* (3rd ed.). Mahwah: Lawrence Erlbaum Associates.
- MacWhinney, B. (2008). How mental models encode embodied linguistic perspectives. In R. Klatzky, B. MacWhinney, & M. Behrmann (Eds.), *Embodiment, ego-space, and action* (pp. 369–410). Mahwah: Lawrence Erlbaum.
- MacWhinney, B. (2011). The logic of the unified model. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition*. New York: Routledge.
- MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: methods for studying discourse. *Aphasiology*, 25, 1286–1307.
- Mantini, D., Hasson, U., Betti, V., Perrucci, M. G., Romani, G. L., et al. (2012). Interspecies activity correlations reveal functional correspondence between monkey and human brain areas. *Nature Methods*, 9, 277–282.
- Marengo, L., Ascoli, G. A., Martone, M. E., Shepherd, G. M., & Miller, P. L. (2008). The NIF LinkOut broker: a web resource to facilitate federated data integration using NCBI identifiers. *Neuroinformatics*, 6, 219–227.
- Marengo, L., Wang, R., Shepherd, G. M., & Miller, P. (2010). The NIF DISCO framework: facilitating automated integration of neuroscience content on the web. *Neuroinformatics*, 8, 101–112.
- Matsumoto, R., Nair, D. R., Bingaman, W., Ikeda, A., Shibasaki, H., & Lüders, H. O. (2005). Intereareal connectivity in the human language system: a cortico-cortical evoked potential study. *International Congress Series*, 1278, 397–400.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48, 67–91.
- McIntosh, A. R. (2012). Tracing the route to path analysis in neuroimaging. *NeuroImage*, 62, 887–890.
- Miller, G. A. (1995). WordNet: a lexical database for english. *Communications of the ACM*, 38, 39–41.
- Nelissen, K., Luppino, G., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2005). Observing others: multiple action representation in the frontal lobe. *Science*, 310, 332–336.
- Nielsen, F. Å. (2005). Mass meta-analysis in Talairach space. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems* (17th ed., pp. 985–992). Cambridge: MIT Press.

- Nielsen, F. Å. (2013). Brede tools and federating online neuroinformatics databases. *Neuroinformatics*. doi:10.1007/s12021-013-9183-4.
- Orban, G. A., Van Essen, D., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences*, 8, 315–324.
- Oztop, E., & Arbib, M. A. (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics*, 87, 116–140.
- Oztop, E., Wolpert, D., & Kawato, M. (2005). Mental state inference using visual control parameters. *Brain Research. Cognitive Brain Research*, 22, 129–151.
- Oztop, E., Kawato, M., & Arbib, M. A. (2013). Mirror neurons: functions, mechanisms and models. *Neuroscience Letters*, 540, 43–55.
- Palmer, M. (2009). Semlink: linking PropBank, VerbNet and FrameNet. *Proceedings of the Generative Lexicon Conference. Sept. 2009, Pisa, Italy: GenLex-09*: 580–590.
- Peeters, R., Simone, L., Nelissen, K., Fabbri-Destro, M., Vanduffel, W., et al. (2009). The representation of tool use in humans and monkeys: common and uniquely human features. *The Journal of Neuroscience*, 29, 11523–39.
- Poizner, H., Klima, E., & Bellugi, U. (1987). *What the hands reveal about the brain*. Cambridge: MIT Press.
- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21, 188–194.
- Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the brain: how our minds share actions, emotions, and experience (Translated from the Italian by Frances Anderson)*. Oxford: Oxford University Press.
- Russ, T. A., Ramakrishnan, C., Hovy, E. H., Bota, M., & Burns, G. A. (2011). Knowledge engineering tools for reasoning with scientific observations and interpretations: a neural connectivity use case. *BMC Bioinformatics*, 12, 351.
- Sagotsky, J. A., Zhang, L., Wang, Z., Martin, S., & Deisboeck, T. S. (2008). Life Sciences and the web: a new era for collaboration. *Molecular Systems Biology*, 4.
- Santos, G. S., Nagasaka, Y., Fujii, N., & Nakahara, H. (2011). Encoding of social state information by neuronal activities in the macaque caudate nucleus. *Social Neuroscience*, 7, 42–58.
- Sato, M. A., Yoshioka, T., Kajihara, S., Toyama, K., Goda, N., et al. (2004). Hierarchical Bayesian estimation for MEG inverse problem. *NeuroImage*, 23, 806–826.
- Silberman, Y., Bentin, S., & Miikkulainen, R. (2007). Semantic boost on episodic associations: an empirically-based computational model. *Cognitive Science*, 31, 645–671.
- Sobolev, A., Leonhardt, A., Rautenberg, P., Kellner, C., Herz, A., Wachtler, T. (2012). Cloud services and a data API for electrophysiology. *Neuroinformatics 2012*: <http://www.neuroinformatics2012.org/abstracts/cloud-services-and-a-data-api-for-electrophysiology>.
- Stein, L. (2002). Creating a bioinformatics nation. *Nature*, 417, 119–120.
- Traxler, M. J., & Gernsbacher, M. A. (Eds.). (2006). *Handbook of psycholinguistics*. London: Academic.
- Turner, J., & Laird, A. R. (2012). The cognitive paradigm ontology: design and application. *Neuroinformatics*, 10, 57–66.
- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92, 231–270.
- Vanduffel, W., Fize, D., Peuskens, H., Denys, K., Sunaert, S., et al. (2002). Extracting 3D from motion: differences in human and monkey intraparietal cortex. *Science*, 298, 413–415.
- Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75, 105–143.
- Vosse, T., & Kempen, G. (2009). The unification space implemented as a localist neural net: predictions and error-tolerance in a constraint-based parser. *Cognitive Neurodynamics*, 3, 331–346.
- Waldrop, M. M. (2008). Science 2.0. *Scientific American*, 298, 68–73.
- Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11, 1317–1329.