NeuroImage 84 (2014) 1061–1069

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Review

# Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data

Florian Hutzler *

*Centre for Neurocognitive Research & Department of Psychology, University of Salzburg, Hellbrunnerstr. 34, 5020 Salzburg, Austria*

ARTICLE INFO

*Article history:*
Accepted 29 December 2012
Available online 11 January 2013

*Keywords:*
Functional specificity
Reverse inference
Cognitive process
Cognition
Brain activation

ABSTRACT

When inferring the presence of a specific cognitive process from observed brain activation a kind of reasoning is applied that is called reverse inference. Poldrack (2006) rightly criticized the careless use of reverse inference. As a consequence, reverse inference is assumed as intrinsically weak by many and its validity has been increasingly regarded as limited. Although it is undisputed that the *careless use* of reverse inference is a problematic practice, the current view of reverse inference is to the author's opinion overly pessimistic. The present manuscript provides a revised formulation of reverse inference that includes an additional conditional constraint that has been previously acknowledged, but so far not implemented: the *task-setting*. This revised formulation I.) reveals that reverse inference can have high predictive power (as demonstrated by an example estimation) and II.) allows an estimation of reverse inference on the basis of meta-analyses instead of large-scale databases. It is concluded that reverse inference cannot be disregarded as a fallacy per se. Rather, the predictive power of reverse inference can even be "decisive"—dependent on the cognitive process of interest, the specific brain region activated, and the task-setting used.



**Contents**



## Introduction

Cognitive science attempts to model human experience and behavior and evidence from functional imaging studies can help to validate these models. One kind of reasoning that is applied is to infer the

* Fax: +43 662 8044 5126.
*E-mail address:* florian.hutzler@sbg.ac.at.

1053-8119 
http://dx.doi.org/10.1016/j.neuroimage.2012.12.075

involvement of a specific cognitive process from observed brain activation during a task. This kind of reasoning is called *reverse inference*.

Whereas the logic of reverse inference is not problematic per se, researchers often neglect the specificity of the activation of a brain region. As pointed out by Poldrack (2006), a specific brain region can be activated by a wide range of cognitive processes. In such a case, it can be problematic to infer the involvement of a specific cognitive process from the activation of this brain region. In other words, the predictive power of reverse inference can be low. Poldrack (2006) addressed this potential fallacy of reverse inference, cautioned against the widespread careless use of reverse inference and warned that researchers should be circumspect in applying this kind of reasoning.

As a consequence, the validity of reverse inference has been increasingly regarded as limited. Brain activation patterns are considered as a weak indicator of the presence of a cognitive process (Poldrack, 2008, 2011) and reverse inference is assumed to be intrinsically weak (Fox and Friston, 2012). Today, researchers applying reverse inference are quickly regarded as falling for "the" fallacy of reverse inference—resulting in the notion of a *general fallacy of reverse inference*.

Although it is undisputed that the careless use of reverse inference is a problematic practice in neuroimaging, the current view of reverse inference is to the author's opinion overly pessimistic. The present manuscript aims to provide a revised formulation of reverse inference that includes an additional (and quite essential) conditional constraint that has been previously acknowledged, but so far not implemented: the *task-setting*.

The revised formulation I.) reveals that reverse inference can have high predictive power (up to the level of being "decisive") and II.) allows an estimation of reverse inference on the basis of meta-analyses instead of large-scale databases. Meta-analyses provide a fine-grained categorization of comparisons. This is an advantage that will become evident when the importance of an adequate classification of comparisons is discussed below.

In the following, the current formulation of reverse inference (as provided by Poldrack, 2006) will be recapitulated. On this basis, the case for a revised formulation will be made. In general, reverse inference allows the determination of the extent to which a certain brain activation is indicative of the involvement of a specific cognitive process (thereafter abbreviated as "the activation" and "the process").

Fig. 1A indicates that the activation can either co-occur with the process or can take place in the absence of the process. Thus, when we back-trace the activation to the level of processes (i.e., when drawing a reverse inference) we can follow two paths. Human reasoning intuitively accounts for the first path: the probability of an activation *in the presence* of the process. We acknowledge the so-called hit-rate when we refer to studies that manipulated the process of interest and when we analyze whether these studies observed the activation in question. The second path is more often neglected: the probability of an activation *in the absence* of the process (i.e., the false-alarm rate). We seldom discuss studies that *did not* manipulate the process of interest but nevertheless resulted in the activation in question.

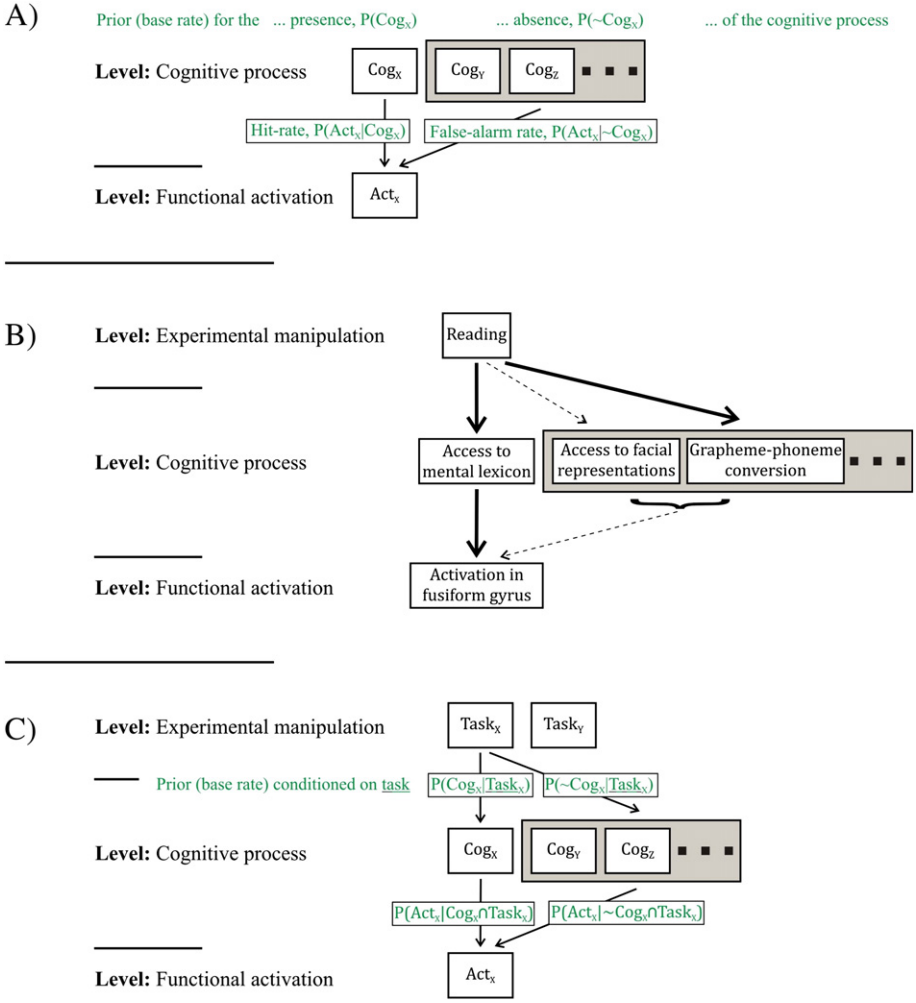


**Fig. 1.** Estimation of reverse inference, schematic depiction of A.) the formulation as realized in Poldrack (2006), B.) an experiment of thought illustrating the necessity for a conditionalization by task, and C.) the revised formulation as used in the present study.

Beneath hit- and false-alarm rates, the probability for the presence of the process itself must also be accounted for—the so-called base rate (see Fig. 1A). The base rate (or probability of a cognitive process being present before observing the activation data) is also known as the *prior*. Below, I will consider prior beliefs about cognitive processes that are conditioned upon the experimental context under which the data were acquired. This can improve the precision of reverse inference; effectively by replacing the prior beliefs (about a cognitive process) with posterior beliefs, given the task or experimental context eliciting those processes.

To summarize, for an estimation of reverse inference the *overall* probability of the activation is necessary in a first step. This overall probability comprises the probability that the activation is observed in the *presence* of the process [P(Act|Cog)P(Cog)] and the probability that the activation is observed in the *absence* of the process [P(Act|~Cog)P(~Cog)]—whereby both of these probabilities are already conditioned on the respective *priors* (i.e., base rates, the underlined part of the formulation). Thus, the overall probability of the activation can be estimated as:

$$P(\mathrm{Act}) = P(\mathrm{Act}|\mathrm{Cog})P(\mathrm{Cog}) + P(\mathrm{Act}|\sim\mathrm{Cog})P(\sim\mathrm{Cog}). \quad (1)$$

In a next step, the probability of the activation in the presence of the process [i.e., the first path; the hit rate conditioned on the *prior*, P(Act|Cog)P(Cog)] is qualified by the overall probability of the activation (as provided in Eq. (1)). The resulting formulation (Eq. (2)) allows us to estimate the predictive quality of reverse inference.

$$P(\mathrm{Cog}|\mathrm{Act}) = \frac{P(\mathrm{Act}|\mathrm{Cog})P(\mathrm{Cog})}{P(\mathrm{Act}|\mathrm{Cog})P(\mathrm{Cog}) + P(\mathrm{Act}|\sim\mathrm{Cog})P(\sim\mathrm{Cog})} \quad (2)$$

### *Bayes factor*

Whereas hit- and false-alarm rates can be determined by appropriate databases, the priors for the presence of the process remain unknown. Poldrack (2006) provided an elegant solution which allows us to circumvent this problem. Poldrack resorted to the basic question, whether the presence of the activation provides additional, substantial evidence for the presence of the process that goes beyond the prior belief about (i.e., the mere base rate of) the process itself.

To do so, Poldrack sets the (unknown) prior belief about the process to an arbitrary constant value (e.g., chance level). The substance of reverse inference can now be formalized by means of the *Bayes factor*: To determine the Bayes factor, the odds [i.e., $P/(1-P)$] for the presence of the process given the activation (i.e., the odds of reverse inference, the numerator in Eq. (3)) are qualified by the odds of the prior belief about the process (i.e., the base rate, the denominator in Eq. (3)).

$$\text{Bayes factor}_{\text{reverse inference}} = \frac{\frac{P(\mathrm{Cog}|\mathrm{Act})}{1-P(\mathrm{Cog}|\mathrm{Act})}}{\frac{P(\mathrm{Cog})}{1-P(\mathrm{Cog})}} \quad (3)$$

The higher the Bayes factor, the greater the predictive quality of reverse inference is. Bayes factors up to 3 are interpreted as "barely worth mentioning" and Bayes factors from 3 to 10 are interpreted as "substantial" (Jeffreys, 1967).

### *Functional specificity is currently assumed to be a prerequisite for reverse inference*

Up to now, a one-to-one correspondence between a brain activation and a cognitive process is assumed to be a prerequisite for reverse inference. The activation of a specific brain region is only regarded as indicative for the presence of a specific cognitive process, if this region is exclusively activated by this very process—but not by alternative processes (as specified by functional specificity, see Cohen and Dehaene, 2004; Dehaene and Cohen, 2011; Kanwisher, 2010).

Functional specificity is integrated into the above formulation of reverse inference by the definition of the *false-alarm rate*: False-alarms are *all* cases during which an activation was observed in the absence of the process—no matter in which experimental context. Such a definition of false-alarms requires considering *every* functional imaging ever being reported—that is, the universe of all comparisons.

Obviously, taking into account *all* comparisons calls for large-scale databases that try to comprise as many comparisons reported in the literature as possible—such as BrainMap (Fox et al., 2005) or NeuroSynth (Yarkoni et al., 2011). Reverse inference in Poldrack (2006), for example, was estimated on the basis of *all* comparisons included in the BrainMap database.

### *The case for a revised formulation*

A thought experiment, however, reveals that reverse inference is not dependent on functional specificity, but rather is dependent on what one might call "task-specific functional specificity". To illustrate, let us assume that an area in the left occipito-temporal region, more specifically, the left fusiform gyrus, is known to be activated by two different processes: a.) access to the mental lexicon (i.e., the recognition of visual words) and b.) face perception.

Let us now assume that a hypothetical experiment during which participants are *presented with visual words* results in an activation of this left fusiform gyrus (see Fig. 1B). Applying the standard formulation of reverse inference, we could now try to resolve, which cognitive processes are likely to co-occur with this activation. We would back-trace from the level of activation to the level of cognitive processes. Doing so would reveal an equally high probability that the observed activation might be indicative for either access to the mental lexicon or for face perception. This finding would imply that activation in the left fusiform gyrus would only provide weak additional evidence for processes related to "visual word recognition", because the observed activation could also be indicative for processes related to "face perception".

Intuitively, however, one would rule out processes of "face perception" and would settle with the conclusion that the observed activation is indicative of "visual word recognition". In what respect does this intuitive reasoning differ from the standard formulation of reverse inference? Intuitively, one would take the *task setting* into account: Being presented with visual words would render processes related to face perception as unlikely to occur.

This additional information about the task being performed allows us to formalize the experimental situation more accurately. Obviously, it improves the precision of reverse inference as it narrows the number of possible alternative explanations (technically: reduces the false alarm rate). Importantly, the necessity of taking the task setting into account has already been acknowledged by Poldrack, stating that "[…] it should be noted that the prior $P(\mathrm{Cog_x})$ is always conditioned on the particular task being used, and should more properly be termed $P(\mathrm{Cog_x}|\mathrm{Task_y})$; however, for the purposes of simplicity I have omitted this additional conditionalization." (Poldrack, 2006, p. 60). In the present manuscript, a revised formulation will be presented that takes into account this important additional constraint.

### *The revised formulation*

Additionally conditioning by task has a massive impact on the formulation and thus the estimation of reverse inference. To illustrate, the level of the task being performed is added in Fig. 1C. Comparing Fig. 1C to A reveals that we are no longer interested in the probability by which the process occurs in any arbitrary situation. Rather, we are interested in the probability by which the process occurs *during a specific task*. More specifically, we are no longer interested in prior beliefs

about the process [i.e., P(Cog)], but in posterior beliefs [prior beliefs that are conditioned upon the experimental context under which the data were acquired, i.e., the base rate conditioned on task, P(Cog|Task)].

Thus, when revising the formulation provided in Eq. (2), the posterior belief is taken into account. The hit rate is now conditioned on the *posterior* [resulting in P(Act|Cog∩Task)P(Cog|Task)]. This conditioned hit rate is then qualified by the overall probability of the activation, whereby now we are interested in the overall probability of the activation *during* the task [i.e., in P(Act|Cog∩Task)P(Cog|Task) + P(Act|~Cog∩Task)P(~Cog|Task)]. The resulting formulation in Eq. (4) now estimates reverse inference while taking into account the task-setting:

$$P(Cog|Act \cap Task) = \frac{P(Act|Cog \cap Task)P(Cog|Task)}{P(Act|Cog \cap Task)P(Cog|Task) + P(Act|\tilde{}Cog \cap Task)P(\tilde{}Cog|Task).} \quad (4)$$

In essence, the formulation of reverse inference in Eq. (4) is analogous to that in Eq. (2), but is additionally conditioned on task. Conditioning by task does *not* affect the estimation of reverse inference via a different probability for the presence of the process. The probability for the presence of the process (whether conditioned on task or not) is regarded as unknown and thus still is set to an arbitrary value (even though the probability of "access to the mental lexicon" might be higher given "being presented with visual words" than without such a task).

Conditioning by task, rather, affects the estimation of reverse inference in an indirect way. Fig. 1B (depicting the above experiment of thought) illustrates that acknowledging for the task "being presented with visual words" reduces the probability that the process "face perception" is among the non-"visual word recognition" process [i.e., to be among ~Cog; whereby it is important to note that P(~Cog) remains unchanged and is set to an arbitrary level]. This reduced probability for "face perception" to be among the non-"visual word recognition" processes lowers the probability for a fusiform activation taking place in the *absence* of visual word recognition processes. In consequence, for our thought experiment, fusiform activation would be quite indicative of visual word recognition. Thus, the precision of reverse inference for revised formulation is increased due to a more accurate task-specific false-alarm rate.

The important consequence for the estimation of reverse inference is that the search space for false alarms is restricted: Of interest are no longer *all* activations of the left fusiform gyrus in the absence of visual word recognition. Rather, of interest are now the activations that occur in the absence of visual word recognition processes, *but in the experimental context of "being presented with visual words"*.

## Example calculation

The current formulation of reverse inference is based on functional specificity and thus needs to account for *all* available comparisons to calculate the false-alarm rate—thus demanding large-scale database such as BrainMap. In contrast, the revised formulation is based on task-specific functional specificity and thus a task-specific false-alarm rate. As a consequence, the revised formulation allows us to resort to a more specialized data basis that is *obtained for a specific task*, such as, e.g., a meta analysis. A meta analysis, being compiled by researchers with expertise in the respective domain, provides a fine-grained cognitive analysis of the processes involved in a comparison. The discussion will address why accurately specifying whether a comparison isolates a specific process or not is *the* prerequisite for a reliable estimation of reverse inference.

### *Data basis*

For the example calculation, a recent meta analysis by Vigneau et al. (2006) was chosen. This meta analysis is dedicated to left hemispheric language processing and reports 729 activation peaks from 260 individual comparisons based on 129 studies. Using this study, the example calculation of reverse inference will be conditioned on the task setting "language processing". Obviously, a more specific (and thus more restrictive) conditioning by, e.g., "visual word recognition" could have been chosen which would have resulted in even higher predictive quality. The lenient conditioning by "language processing", however, is sufficient for this exploratory estimation.

Importantly, Vigneau et al. only included comparisons which aimed at isolating a specific cognitive process, whereas low-level contrasts (e.g., against baseline) were not included. To illustrate, a comparison classified as, e.g., phonological comparison specifically targeted phonological processing but did not isolate, e.g., semantic processing. Three different classes of cognitive processes were defined in Vigneau et al.: phonological, semantic, and sentence processing.

### *Analysis*

The 729 activation peaks provided in Vigneau et al.'s supplementary material were aggregated to the level of the 260 comparisons. For estimation, a sphere with a predefined radius of 10 voxels around the coordinate of interest was searched for the presence of activation peaks. Upon these activation peaks, the hit-rates (i.e., based on activation peaks observed during the process of interest) and false-alarm rates (i.e., based on activation peaks observed during the absence of the process of interest) are calculated—whereby these probabilities are estimated on the level of comparisons, not individual activation peaks. These rates are automatically conditioned on task due to the task-specific data basis chosen. Importantly, since the task-setting refers to the subset of tasks included in Vigneau et al., the task variable was constant throughout the analyses. On the basis of these probabilities, reverse inference was calculated along the revised formulation provided in Eq. (4). A calculation was only performed when a minimum of 5 activation peaks was located within the sphere. If fewer peaks were present, the Bayes factor for this specific coordinate was determined as "not supporting" the presence of a process (i.e., Bayes factor 0).

### *Brain maps of reverse inference*

For the example calculation, the Bayes factor was calculated analog to Eq. (3)'s logic. However, since the revised formulation is conditioned upon the experimental context, the prior was replaced by the posterior (i.e., the task-specific base rate). Thus, the odds of reverse inference as estimated by the revised formulation (the numerator in Eq. (5)) are qualified by the odds of the posterior (the denominator in Eq. (5)). Appendix A provides a worked example of the example calculation for a single voxel.

$$\text{Bayes factor}_{\text{revised reverse inference}} = \frac{\frac{P(Cog|Act \cap Task)}{1-P(Cog|Act \cap Task)}}{\frac{P(Cog|Task)}{1-P(Cog|Task)}} \quad (5)$$

The Bayes factor was calculated for every voxel in the brain and brain maps of reverse inference were generated for demonstrative purposes. These brain maps provide an example atlas of the predictive quality of reverse inference for a specific cognitive process (as classified by Vigneau et al.), revealing, for which coordinates a reliable reverse inference is possible, and for which not.

### *Results*

For visualization of reverse inference, only Bayes factors which provide "strong" evidence (i.e., >= 10) are indicated. Because of the demonstrative purpose of the estimation, brain maps are only presented cursorily. Vigneau et al.'s meta analysis only served as a test-bed and future studies might resort to even more specialized meta-analyses. Furthermore, it is important to note that Vigneau et al.'s meta analysis

(and thus the present study's estimation of reverse inference) was limited to left hemispheric language processing (Figs. 2–4).

## Discussion

### *Reverse inference is not intrinsically weak*

Currently, reverse inference is considered as *generally* being of poor predictive power and as being intrinsically weak. An example calculation on the basis of the revised formulation, however, revealed a more differentiated pattern of results: Activations in certain brain regions can provide even "decisive" evidence for the presence of a specific cognitive process. For other areas, no such reverse inference can be drawn. Thus, the present study suggests that reverse inference is not a fallacy per se. Rather, its reliability is dependent on the cognitive process of interest, the task-setting used—and has to be individually estimated for the specific brain activation in question.

### *Reverse inference does not depend on functional specificity*

Whether the logic of reverse inference is feasible for the interpretation of a study depends on our intended interpretive direction: Do we

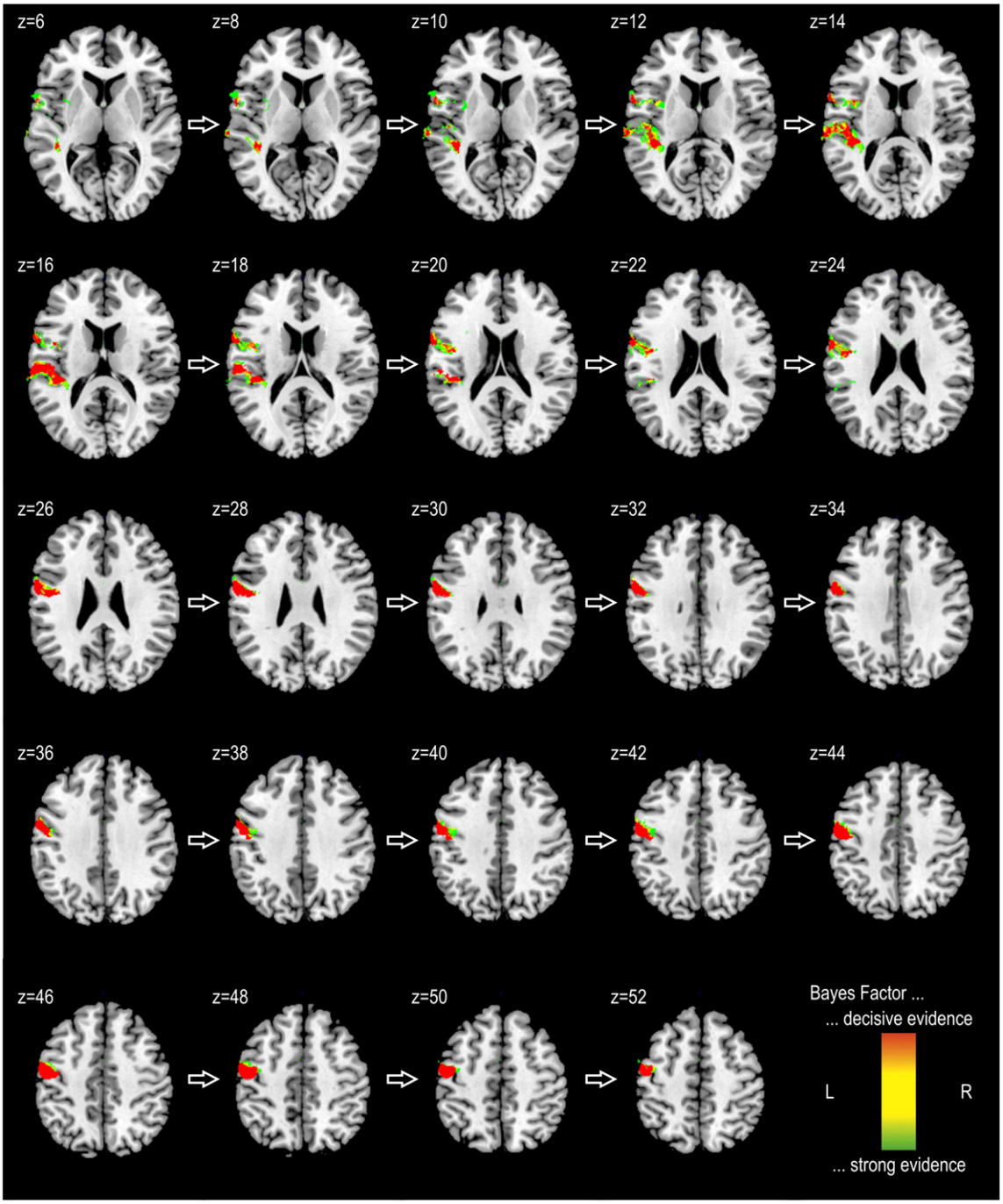


**Fig. 2.** Brain maps indicating the predictive power of reverse inference for phonological processing. Two coherent clusters could be identified in the left hemisphere. A first cluster is located in the left precentral region, elongating dorsoventrally from z = 10 to z = 54. A second cluster is located in the perisylvian region, elongating dorsoventrally from z = 3 to z = 23. The strength of reverse inference is "decisive" (i.e., a Bayes factors >= 20) for the majority of the voxels in both clusters.

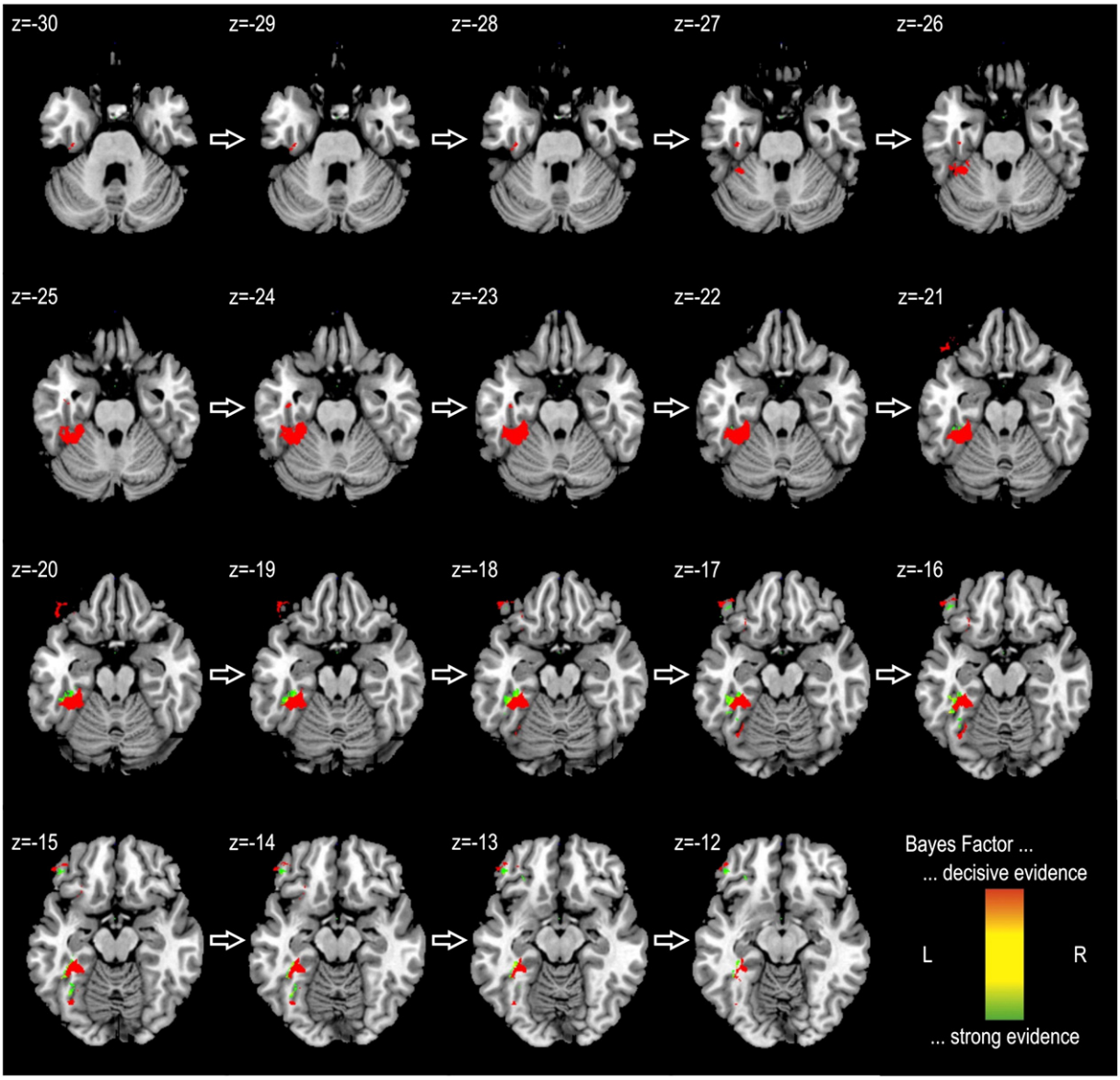


**Fig. 3.** Brain maps indicating the predictive power of reverse inference for semantic processing. A single cluster of voxels (providing "decisive" evidence) in the area around the fusiform gyrus was found to be indicative for semantic processing.

want to use functional imaging (as one source of evidence among others) to decide among competing theories for human behavior? If so, then we want to know, whether a specific brain activation assessed under specific circumstances can be *indicative* for the presence of a certain cognitive process. More specifically, we want to know whether *during a specific task*, an activation is indicative for a specific cognitive process. For such reasoning it is not problematic when this very brain region is also activated by other cognitive processes *during a different task*. To illustrate, an area activated by the process "visual word recognition" and the process "face recognition" can be indicative for "visual word recognition" as long as the stimuli are words but not faces.

Taken together, the differences between the current and the revised formulation of reverse inference can be boiled down to one core issue: Functional specificity is a prerequisite for the *current* formulation of reverse inference. The fact that the "same [brain] structure can be assigned very different functions" (Poldrack, 2012) was regarded as *the* fundamental problem for reverse inference. In contrast, the *revised* formulation is conditioned on the task-setting used and, in consequence, is no more dependent on a one-to-one correspondence between a brain activation and a cognitive process.

The necessity of conditioning by task has already been acknowledged by Poldrack (2006, p. 60), but not implemented up to now. To this end, the revised formulation is the logical further development of the current formulation. The revised formulation allows us to formalize the experimental context more accurately and results in a considerably increased precision of reverse inference. Conditioning by task therefore seems not only optional, but even mandatory: Every piece of information that helps to increase the predictive *quality* of reverse inference needs to be accounted for—and the task-setting is such a quintessential piece of information.

### *For reverse inference, large-scale databases are not mandatory*

The formulation of reverse inference does not only determine its predictive power. Rather, it also determines upon which kind of data-basis reverse inference can be estimated. The current formulation of reverse inference requires the inclusion of as many comparisons as possible, with the ultimate goal of accounting for every valid comparison ever being published. Obviously, large-scale databases

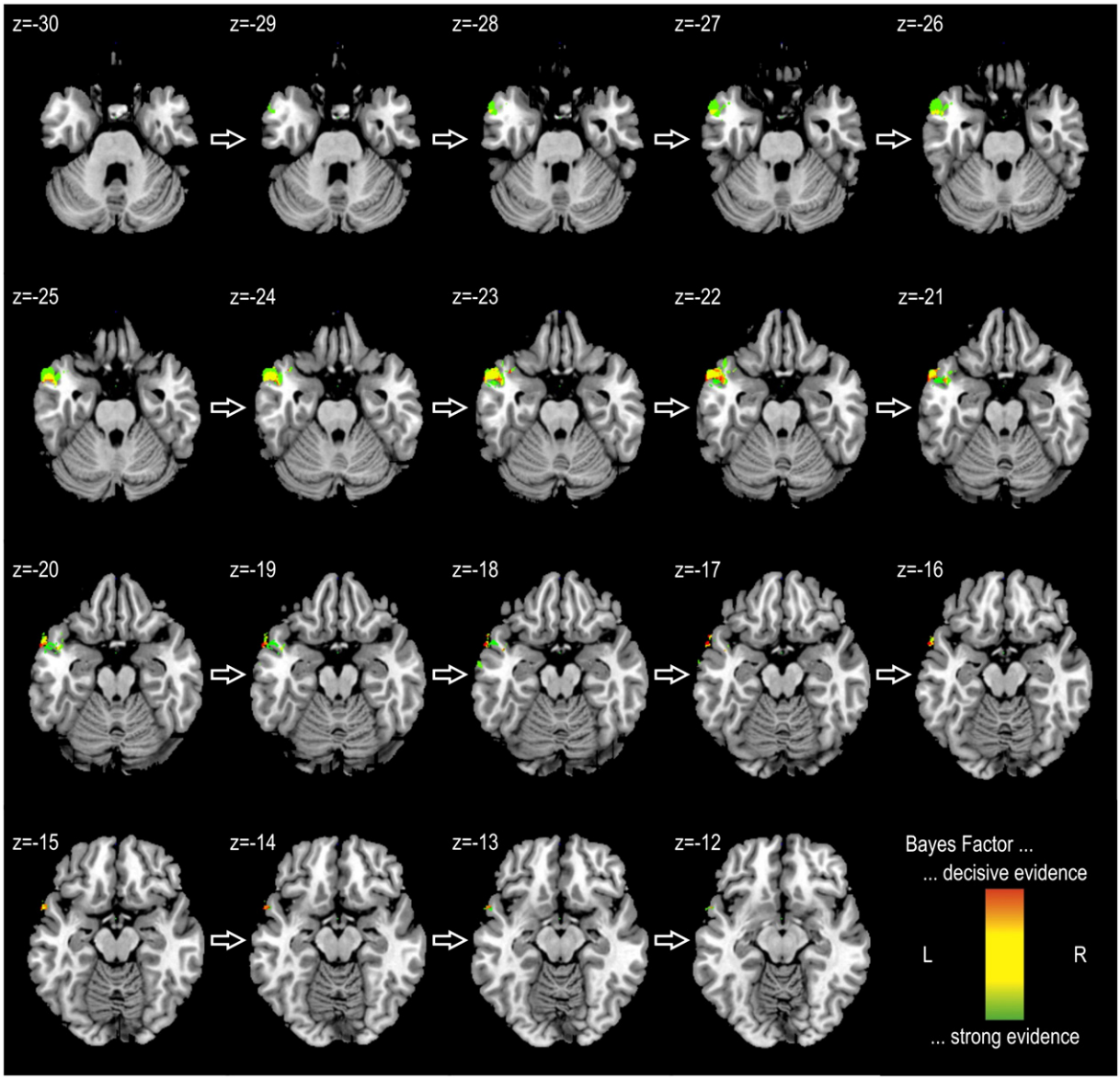


**Fig. 4.** Brain maps indicating the predictive power of reverse inference for sentence processing. A small cluster providing "strong", but not "decisive", evidence was found in the temporal pole.

such as BrainMap (Fox et al., 2005) or NeuroSynth (Yarkoni et al., 2011) are necessary for such an approach.

Estimating reverse inference on the basis of large-scale databases, however, is error-prone. The key to a reliable calculation of reverse inference is an adequate classification of comparisons. On the one hand, we need to identify the comparisons that *isolate* a specific process, thus constituting the hit-rate. On the other hand, we need to identify those comparisons that *definitely do not* isolate this process, thus constituting the false-alarm rate. Given the variety of alternative processes that can be isolated by the comparisons that comprise the false-alarm rate, it is easily overseen that this second group of comparisons *must under no circumstance* isolate the process of interest.

Such an adequate classification is not always possible on the basis of large-scale databases. This problem can be exemplified by BrainMap's behavioral domain code, which was used in previous estimations of reverse inference. This code classifies the mental operations (cognitive processes among other operations) that are intended to be isolated by a statistical contrast (Laird et al., 2005). Importantly, there is no exclusive mapping between the code and a specific cognitive process. Rather, the code specifies whether a comparison was intended to isolate *one or many processes* that belong to a certain domain, irrespective of whether such a comparison might *also isolate other processes* that are related to *other* domains. Thus, a code comprises a heterogeneous compilation of comparisons that differ with respect to the type and sheer number of processes being isolated.

To illustrate, concerning the hit-rate: a comparison classified by the code X does not necessarily isolate *only* the process $Cog_x$, but might also isolate processes like $Cog_y$ that result in unforeseeable additional activations. Concerning the false-alarm rate, a comparison classified by the domain code Y might (beneath the intended process $Cog_y$) *also* isolate an uncalled-for process $Cog_x$—with the foreseeable, additional activation $Act_x$ that would misleadingly be rated as a false-alarm.

Thus, BrainMap's behavioral domain code (originally not intended to fulfill the specific requirements for reverse inference) is not adequate to estimate reverse inference. When one considers the universe of all existing comparisons and the number of cognitive processes that are potentially of interest, then the dimension (not to say impossibility) of a universal classification system that is suitable for reverse inference becomes evident.

*Reverse inference can be estimated on the basis of meta-analyses*

In contrast, the revised formulation (being conditioned on task) allows an estimation of reverse inference on the basis of specialized meta-analyses. Meta analyses provide the advantage of a fine-grained categorization of comparisons on the basis of a cognitive analysis done by experts in the respective domain. This categorization is a prerequisite for a reliable estimation of reverse inference. Thus, additionally conditioning by task relieves us from going for mass, but allows us to go for detail.

## Limitation of the example calculation

Importantly, the present manuscript's example calculation is only for demonstrative purpose: Vigneau et al.'s (2006) meta-analysis only served as a test bed for the revised formulation. The categorization used is rather coarse, distinguishing between phonological, semantic, and sentence processing. Moreover, the task-setting used for conditioning (i.e., language-processing) is quite broad. Future studies might want to resort to more specialized meta-analyses that narrow the conditioned task setting and provide a more restricted categorization of comparisons. Obviously, such a more specialized data-basis will further improve the predictive quality of reverse inference.

The revised formulation, however, also implicates an important constraint: When reasoning along reverse inference, no cross-task inferences are admissible since the task-setting is *the* additional conditional constraint.

Up to now, the discussion on reverse inference focused on the specificity of an activation, i.e., the question whether reverse inference is possible if a specific brain region is activated by more than one cognitive process. The effect of cognitive degeneracy (Price and Friston, 2002), however, was not acknowledged yet: a structure–function relationship in which more than one neuronal system produces the same response. Most probably, the impact of cognitive degeneracy on the predictive quality of reverse inference might depend on the grain-size of the cognitive process of interest (e.g., "reading" or more specifically "grapheme–phoneme conversion"). This question demands a closer examination which is beyond the scope of this manuscript.

## Acknowledgments

The author would like to thank Andreas Huttegger for computationally implementing the estimations of reverse inference as well as the anonymous reviewers and the editors for the valuable comments that greatly helped to improve the manuscript.

## Appendix A. Revised formulation of reverse inference: worked example of the calculation

The present manuscript's example calculation is based on a meta-analysis performed by Vigneau et al. (2006). In the supplementary material, Vigneau et al. (2006) report 729 activation peaks, whereby 247 of these peaks correspond to phonological processing and 482 peaks correspond to non-phonological processing. These peaks result from 260 comparisons, whereby 74 of these comparisons were performed to isolate phonological processes and 186 of these comparisons were performed to isolate non-phonological processes. It is important to note that the probabilities used for the estimation of reverse inference are estimated on the level of comparisons, but not on the level of individual activation peaks. For the worked example, let us consider the coordinate −60,−17,16. Within a sphere of 10 voxels around (i.e., Euclidean distance to) this coordinate, activation peaks from 8 different comparisons can be found. 7 of these comparisons targeted phonological processing, 1 of these comparisons targeted non-phonological processing.

On the basis of these numbers, we can now estimate the following probabilities:

- The task-specific hit rate, P(Act|Cog ∩ Task): During the presence of phonological processes, an activation within this sphere could be observed in 7 out of 74 comparisons, i.e., P(Act|Cog ∩ Task) = 7/74 (≈.0946)
- The task-specific false-alarm rate, P(Act|~Cog ∩ Task): During the absence of phonological processing, an activation in this sphere could be observed in 1 out of 186 comparisons, i.e., P(Act|~Cog ∩ Task) = 1/186 (≈.0054)
- The posterior beliefs about cognitive processes (i.e., prior beliefs about cognitive processes conditioned upon the experimental context or: the task-specific base-rate) are unknown and thus are set to an arbitrary level, P(Cog|Task) = .5 and P(~Cog|Task) = .5, respectively.

We can now resort to Eq. (4) (the revised formulation of reverse inference) to estimate the predictive quality of reverse inference:

$$P(Cog|Act\cap Task)=\frac{P(Act|Cog\cap Task)P(Cog|Task)}{P(Act|Cog\cap Task)P(Cog|Task)+P(Act|\sim Cog\cap Task)P(\sim Cog|Task)}=\frac{{}^{7}\!/_{74}\times .5}{{}^{7}\!/_{74}\times .5+{}^{1}\!/_{186}\times .5}\approx .95$$

Thus, the predictive quality of reverse inference is .95. In other words, when we observe an activation at the example coordinate we can (under the given experimental conditions) conclude with a probability of .95 upon the presence of a phonological process. Using Eq. (5), we can now estimate, whether the presence of an activation provides additional, substantial evidence for the presence of a phonological process that goes beyond the mere posterior belief about the process (i.e., the task-specific base-rate):

$$\text{Bayes factor}_{\text{revised reverse inference}}=\frac{P(Cog|Act\cap Task)}{1-P(Cog|Act\cap Task)}\Big/\frac{P(Cog|Task)}{1-P(Cog|Task)}=\frac{.95}{1-.95}\Big/\frac{.5}{1-.5}=17.6$$

A Bayes factor of 17.6 reveals that observing an activation at this coordinate provides "strong evidence" for the presence of a phonological process that goes beyond the mere posterior belief about the process itself. In consequence, in the brain maps indicating predictive power of reverse inference for phonological processing, the voxel at −60,−17,16 is set to a value of 17.6.

## References

Cohen, L., Dehaene, S., 2004. Specialization within the ventral stream: the case for the visual word form area. NeuroImage 22 (1), 466–476.
Dehaene, S., Cohen, L., 2011. The unique role of the visual word form area in reading. Trends Cogn. Sci. 15 (6), 254–262.
Fox, P.T., Friston, K.J., 2012. Distributed processing; distributed functions? NeuroImage 61 (2), 407–426.
Fox, P.T., Laird, A.R., Fox, S.P., Fox, P.M., Uecker, A.M., Crank, M., et al., 2005. BrainMap taxonomy of experimental design: description and evaluation. Hum. Brain Mapp. 25 (1), 185–198.
Jeffreys, L., 1967. Theory of Probability. Clarendon Press, Oxford.
Kanwisher, N., 2010. Functional specificity in the human brain: a window into the functional architecture of the mind. Proc. Natl. Acad. Sci. 107 (25), 11163–11170.
Laird, A.R., Lancaster, J.L., Fox, P.T., 2005. BrainMap—the social evolution of a human brain mapping database. Neuroinformatics 3 (1), 65–77.
Poldrack, R.A., 2006. Can cognitive processes be inferred from neuroimaging data? Trends Cogn. Sci. 10 (2), 59–63.
Poldrack, R.A., 2008. The role of fMRI in Cognitive Neuroscience: where do we stand? Curr. Opin. Neurobiol. 18 (2), 223–226.

Poldrack, R.A., 2011. Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. Neuron 72 (5), 692–697.
Poldrack, R.A., 2012. The future of fMRI in Cognitive Neuroscience. NeuroImage 62 (2), 1216–1220.
Price, C.J., Friston, K.J., 2002. Degeneracy and cognitive anatomy. Trends Cogn. Sci. 6 (10), 416–421.
Vigneau, M., Beaucousin, V., Herve, P.Y., Duffau, H., Crivello, F., Houde, O., et al., 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. NeuroImage 30 (4), 1414–1432.
Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8 (8), 665–670.