NeuroImage 59 (2012) 1846–1858

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

# Componential Granger causality, and its application to identifying the source and mechanisms of the top–down biased activation that controls attention to affective vs sensory processing

Tian Ge [a], Jianfeng Feng [a,b,**], Fabian Grabenhorst [c], Edmund T. Rolls [b,d,*]

[a] Centre for Computational Systems Biology, School of Mathematical Sciences, Fudan University, Shanghai, China
[b] Department of Computer Science, University of Warwick, CV4 7AL, UK
[c] University of Cambridge, Department of Physiology, Development and Neuroscience, Cambridge, UK
[d] Oxford Centre for Computational Neuroscience, Oxford, UK

ARTICLE INFO

*Article history:*
Received 19 May 2011
Revised 13 July 2011
Accepted 15 August 2011
Available online 23 August 2011

*Keywords:*
Selective attention
Granger causality
Causal networks
Functional networks
Biased competition
Biased activation theory of attention
Value
Taste
Affect
Emotion
Connectivity
Primary taste cortex
Orbitofrontal cortex
Dorsolateral prefrontal cortex
Reward
Pleasantness
Intensity
Neuroimaging
fMRI
PPI

ABSTRACT

We describe a new measure of Granger causality, componential Granger causality, and show how it can be applied to the identification of the directionality of influences between brain areas with functional neuroimaging data. Componential Granger causality measures the effect of $y$ on $x$, but allows interaction effects between $y$ and $x$ to be measured. In addition, the terms in componential Granger causality sum to 1, allowing causal effects to be directly compared between systems. We show using componential Granger causality analysis applied to an fMRI investigation that there is a top–down attentional effect from the anterior dorsolateral prefrontal cortex to the orbitofrontal cortex when attention is paid to the pleasantness of a taste, and that this effect depends on the activity in the orbitofrontal cortex as shown by the interaction term. Correspondingly there is a top–down attentional effect from the posterior dorsolateral prefrontal cortex to the insular primary taste cortex when attention is paid to the intensity of a taste, and this effect depends on the activity of the insular primary taste cortex as shown by the interaction term. Componential Granger causality thus not only can reveal the directionality of effects between areas (and these can be bidirectional), but also allows the mechanisms to be understood in terms of whether the causal influence of one system on another depends on the state of the system being causally influenced. Componential Granger causality measures the full effects of second order statistics by including variance and covariance effects between each time series, thus allowing interaction effects to be measured, and also provides a systematic framework within which to measure the effects of cross, self, and noise contributions to causality. The findings reveal some of the mechanisms involved in a biased activation theory of selective attention.



## Introduction

Correlations between signals, including signals at the neuronal or at the functional neuroimaging level, do not reveal the direction of the possible influence of one signal on the other. Understanding how one brain area may influence another, for example by providing it with inputs, or by top–down modulation, is fundamental to understanding how the brain functions (Bar, 2007; Bressler and Menon, 2010; Mechelli et al., 2004). Hence, inferring causal influences from time series data has been attracting intensive interest. Recently, Granger causality has become increasingly popular due to its easy implementation and many successful applications to econometrics, neuroscience, etc., and in particular, the study of brain function (Bressler et al., 2008; Bressler and Seth, 2011; Deshpande et al., 2010a; Ding et al., 2006; Hwang et al., 2010; Jiao et al., 2011; Luo et al., 2011; Schippers et al., 2010).

Granger causality is based on precedence and predictability. Originally proposed by Wiener (1956) and further formalized by Granger (1969), it states that given two times series $x$ and $y$, if the inclusion of the past history of $y$ helps to predict the future states of $x$ in some

* Correspondence to: E.T. Rolls, Oxford Centre for Computational Neuroscience, Oxford, UK.
** Correspondence to: J. Feng, Department of Computer Science, University of Warwick, Coventry CV4 7AL.
*E-mail addresses:* jianfeng.feng@warwick.ac.uk (J. Feng), Edmund.Rolls@oxcns.org (E.T. Rolls).
*URL:* http://www.oxcns.org (E.T. Rolls).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.08.047

plausible statistical sense, then $y$ is a cause of $x$ in the Granger sense. In spite of the wide acceptance of this definition, we show in this paper that classical Granger causality is not tailored to measure the effects of interactions between time series $x$ and $y$ on the causal influences, and cannot measure systematically the effects of the past history of $x$ on $x$. A componential form of Granger causality analysis is therefore introduced and described here, which has advantages over classical Granger analysis. Specifically, componential Granger causality extends classical Granger causality by taking account of both the variance and the interaction terms of second order statistics, and is a more systematic way to infer causal influences from various sources, which allows relative effects within a system to be measured. Componential Granger causality thus usefully extends and complements classical Granger causality.

The componential Granger analysis is then applied to an fMRI investigation to identify the source of the top–down selective attentional control that has been shown to differentially bias brain systems involved in affective vs sensory analysis (Grabenhorst and Rolls, 2008, 2011; Rolls et al., 2008). Instructions to pay attention to and later rate the pleasantness of a taste increase the activations to taste measured with functional magnetic resonance imaging (fMRI) in the orbitofrontal and pregenual cingulate cortices (Grabenhorst and Rolls, 2008), where the subjective pleasantness of taste is represented (de Araujo et al., 2003b; Grabenhorst et al., 2008, 2010; Kringelbach et al., 2003; Rolls and Grabenhorst, 2008), but not the primary taste cortex in the anterior insula (Grabenhorst and Rolls, 2008), where the subjective intensity and identity of taste are represented (de Araujo et al., 2003b; Grabenhorst et al., 2008; Haase et al., 2009; Kringelbach et al., 2003; Rolls and Grabenhorst, 2008; Small et al., 2003). Instructions to pay attention to and later rate the intensity of a taste increase the activations to taste in the insular taste cortex but not in the orbitofrontal and pregenual cingulate cortices (Grabenhorst and Rolls, 2008). This differential biasing of brain systems for affect-related vs more sensory-related processing may be an important aspect of how top–down cognition and selective attention operate (Grabenhorst and Rolls, 2011).

Top–down selective attention involves a modulation, probably by biased competition, of the responses of neurons to incoming sensory stimuli (Deco and Rolls, 2005a; Desimone and Duncan, 1995; Rolls, 2008; Rolls and Deco, 2002). Here we investigate the source of the top–down attentional modulation, thought often to be some part of the prefrontal cortex (Corbetta and Shulman, 2002; Downar et al., 2000; Kanwisher and Wojciulik, 2000), where a short-term memory can keep active the neurons representing the attentional task instruction or rule to provide the source of the top–down bias. Using the psychophysiological interaction (PPI) connectivity approach (Friston et al., 1997; Gitelman et al., 2003), we identified an anterior lateral prefrontal cortex region at Y = 53 mm in which the correlation with activity in the orbitofrontal cortex seed region was greater when attention was to pleasantness than to intensity (Grabenhorst and Rolls, 2010). Conversely, in a more posterior region of lateral prefrontal cortex at Y = 34 the correlation with activity in the anterior insula seed region was greater when attention was to intensity than to pleasantness.

## Methods

The PPI analyses described previously (Grabenhorst and Rolls, 2010) showed in a task on which attention was directed either toward the pleasantness of a taste or its intensity, that an anterior lateral prefrontal cortex region at Y = 53 mm had a correlation with activity in the orbitofrontal cortex seed region that was greater when attention was to pleasantness than to intensity (Grabenhorst and Rolls, 2010). Conversely, in a more posterior region of lateral prefrontal cortex at Y = 34 the correlation with activity in the anterior insula seed region was greater when attention was to intensity than to pleasantness. However, such PPI analyses do not show the directionality of the influences, as they are based on correlations. Granger causality analysis in principle enables one to assess the directionality of the influences (Ding et al., 2006; Guo et al., 2010). Classical Granger causality analysis assesses the effect of the past history of $y$ on $x$, relative to the noise (where the noise is the variance of the error in predicting $x$ from $y$). We show below that a problem with this classical Granger causality is that the measure is not comparable between different systems. Here we introduce a new approach, componential Granger causality, in which there are four components that sum to 1, allowing causality in different systems to be directly compared (because they are on the same scale). Moreover, the nature of the causality can be interpreted, by assessing the magnitudes of the four components, namely an auto-component (an effect of the past history of $x$ on $x$); a cross component (an effect of the past history of $y$ on $x$); an interaction term (an effect of the joint past history of $x$ and $y$ (their covariance) on $x$); and the noise defined above. We then apply the componential Granger causality analysis to an fMRI study, and show how it can be used to interpret the effects of one brain area on another, and vice versa, and whether these effects involve interactions between the activities in the two brain areas.

### *Componential Granger causality analysis*

The past few years have witnessed a significant growth in the application of Granger causality to various research areas, especially to neurophysiological recordings and functional MRI data. Proposed by Granger (1969) and further extended to the frequency domain by Geweke (Geweke, 1982; Geweke, 1984), Granger causality has become a powerful tool to detect causal influences and functional connectivity in temporal data from various sources (Ding et al., 2006; Green, 2002; Sims, 1972).

Suppose the vector stochastic process $x_t$ is stationary and admits the vector autoregressive representation $x_t = A(\mathcal{L})x_{t-1} + \varepsilon_t$ where $x_t = [x_{1t}, x_{2t}, \cdots, x_{Nt}]^T$, $A(\mathcal{L}) = \left[a_{ij}(\mathcal{L})\right]_{N\times N}$, $a_{ij}(\mathcal{L}) = \sum_{k=0}^{+\infty} a_{ij,k}\mathcal{L}^k$, $\mathcal{L}$ is the lag operator satisfying $\mathcal{L}x_t = x_{t-1}$, $\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}, \cdots, \varepsilon_{Nt}]^T$ is white noise, and $[\cdot]^T$ is the transpose of a vector.

We firstly define classical Granger causality from $x_{it}$ to $x_{jt}$ as $C_{i\to j} = \ln\{1 + \mathrm{var}[a_{ji}(\mathcal{L})x_{i,t-1}] / \mathrm{var}[\varepsilon_{jt}]\}$, where var[·] is the variance of a random variable (Ding et al., 2006). Classical Granger causality thus measures for example whether previous values of $x_i$ help in the prediction of $x_j$, and scales this relative to the variance of the noise. Classical Granger causality can also be expressed as reducing the prediction error of $x_j$ with and without the previous history of $x_i$ and this can be affected by the covariance between $x_i$ and other terms, though it does not explicitly measure and separate out effects of the covariance terms (Ding et al., 2006; Granger, 1969).

In spite of its easy implementation, wide justification and many successful applications, the classical definition of Granger causality is problematic when used to measure causal strength, for a number of reasons. First, the magnitude is unbounded, as it depends on the effect of $y$ on $x$, divided by the noise (where the noise is the variance of the error in predicting $x$ from $y$). If an alternative measure was on the same scale of 0 to 1 for all systems (e.g. that all the components sum to 1), that might be advantageous. Second, classical Granger causality does not take into account the effect of the past history of $x$ on $x$, and the magnitude of this compared to the effect of the past history of $y$ on $x$ may well be of interest. Third, the classical measure takes no account of any covariance between $x$ and $y$, which may be crucial in interpreting the causality. For example, one brain region may influence a second brain region only when the second region is responding, and that interaction effect is not captured by classical Granger causality.

To overcome these shortcomings, we propose a novel definition, componential Granger causality, to measure causal strength. We first describe the componential Granger causality measure with a simple model with a time series $x$ that may be influenced by a time series $y$. Then we generalize this to many time series.

The two-dimensional model has two variables $x$ and $y$:

$$\begin{cases} x_t = a_{xx}(\mathcal{L})x_{t-1} + a_{yx}(\mathcal{L})y_{t-1} + \varepsilon_{xt} \\ y_t = a_{xy}(\mathcal{L})x_{t-1} + a_{yy}(\mathcal{L})y_{t-1} + \varepsilon_{yt} \end{cases}. \quad (1)$$

The components that we define for componential Granger causality are:

Component 1: The effect of the past history of $y$ on $x$:

$$F_{y\to x} = \frac{\mathrm{var}\left[a_{yx}(\mathcal{L})y_{t-1}\right]}{\mathrm{var}[x_t]}. \quad (2)$$

Component 2: The effect of the past history of $x$ on $x$:

$$F_{x\to x} = \frac{\mathrm{var}[a_{xx}(\mathcal{L})x_{t-1}]}{\mathrm{var}[x_t]}. \quad (3)$$

Component 3: The effect of the interaction between $x$ and $y$ on $x$:

$$F_{(x,y)\to x} = \frac{2\,\mathrm{cov}\left[a_{xx}(\mathcal{L})x_{t-1}, a_{yx}(\mathcal{L})y_{t-1}\right]}{\mathrm{var}[x_t]}. \quad (4)$$

Note that $F_{(x,y)\to x}$ is in general different from $F_{(x,y)\to y}$.

Component 4: The effect of the noise on $x$:

$$F_{noise_x\to x} = \frac{\mathrm{var}[\varepsilon_{xt}]}{\mathrm{var}[x_t]}. \quad (5)$$

The sum of the above four components is 1.

Componential Granger causality is defined by the following three terms, which sum to 1:

$$F^c_{y\to x} = F_{y\to x} + \frac{1}{2}F_{(x,y)\to x}. \quad (6)$$

(which we term cross-componential Granger causality). (This is the componential Granger causality equivalent of classical Granger causality, and might be referred to as componential Granger causality in the context of only effects of $y$ on $x$ being considered.)

$$F^c_{x\to x} = F_{x\to x} + \frac{1}{2}F_{(x,y)\to x}. \quad (7)$$

(which we term auto-componential Granger causality.)

$$F^c_{noise_x\to x} = F_{noise_x\to x}. \quad (8)$$

(This can be thought of as the noise or otherwise unattributed effects on $x$.)

The cross term of componential Granger causality $F^c_{y\to x}$ defined in Eq. (6) is Component 1 defined in Eq. (2) (which corresponds to classical Granger causality), plus half of the covariance component defined as Component 3. Thus a difference from classical Granger causality is that the cross term of componential Granger causality as defined in Eq. (6) also takes into account the interaction between the signal term $y$ and its 'carrier' $x$. Intuitively, a positive interaction between $x$ and $y$ increases the cross term of componential Granger causality, and a negative interaction between $x$ and $y$ reduces the cross term of componential Granger causality. This is one of the main reasons we introduce componential Granger causality. Classical Granger causality does not measure the interaction term (Ding et al., 2006; Green, 2002; Sims, 1972), and that is the most important innovation that we introduce in this paper. We note that classical Granger causality measures only a part of second order statistics, the variance, and ignores the interaction (or cross-correlation) effects. Componential Granger causality extends classical Granger causality by taking account of both the variance and the interaction terms of second order statistics. The cross term of componential Granger causality $F^c_{y\to x}$ defined in Eq. (6) can become negative if half of Component 3 when negative is larger than Component 1. The auto or self term defined in Eq. (7) is Component 2 plus half of the covariance component defined as Component 3. The third term defined in Eq. (8) is the effect of the noise on $x$ defined in Component 4. Componential Granger causality as defined here has the property that its three terms (Eqs. (6)–(8)) sum to one. This allows the *relative magnitude* due to an effect of $y$ on $x$, the effect of $x$ on $x$, with their interactions, to be compared between systems (because the terms in Eqs. (6)–(8) sum to 1). Classical Granger causality (see definition above) measures effects of the past history of $y$ on $x$, and expresses this relative to the variance of the noise, and so enables the *absolute magnitude* of causal effects to be compared between systems. This is a useful feature of classical Granger causality that allows absolute effects to be measured, and complements the useful feature of componential Granger causality that allows relative effects within a system to be measured. However, in componential Granger causality, absolute effects are represented by how large the contributions of the cross-componential (Eq. (6)) and autocomponential (Eq. (7)) terms are compared to the noise term (Eq. (8)), given that these three terms sum to 1.

Basically, the variance of the target time series $x_t$ is decomposed into the driving forces from various sources including itself and the noise. The covariance term $\mathrm{cov}\left[a_{xx}(\mathcal{L})x_{t-1}, a_{yx}(\mathcal{L})y_{t-1}\right]/\mathrm{var}[x_t]$ in this definition can be interpreted as the causal strength from the interaction between $x$ and $y$ on $x$. A positive value implies that including the past history of $y$ improves the prediction of the current state of $x$, reflecting a positive correlation between $x$ and $y$. A negative value implies that including the history of $y$ also improves the prediction of the current state of $x$, but reflects a negative correlation between $x$ and $y$. A value of zero for the covariance term indicates there is no causal influence from the interaction between $x$ and $y$ on $x$. This could arise either when the inclusion of the history of $x$ or $y$ will not help predict the current of $x$, or when the history of $x$ and $y$ are independent of each other, i.e., there is no interaction between $x$ and $y$. Therefore, different driving forces might either cooperate to enhance the causal influences or compete and counteract with each other. Sometimes, the negative effect of the interaction between $x$ and $y$ can exceed the contribution of the history of $y$. This corresponds to the situation in Eq. (6) that $\mathrm{var}\left[a_{yx}(\mathcal{L})y_{t-1}\right] < -\mathrm{cov}\left[a_{xx}(\mathcal{L})x_{t-1}, a_{yx}(\mathcal{L})y_{t-1}\right]$, which leads to a negative componential Granger causality from $y$ to $x$.

We note in addition that the ratio of Component 1 to Component 2 measures the contribution of the past history of $y$ on $x$ relative to the past history of $x$ on $x$.

Further reasons for the definition that we provide for componential Granger causality, and some of its advantages, are provided in Appendix A.

To generalize to the case with multiple time series, suppose the vector stochastic process $x_t$ is stationary and admits the vector autoregressive representation $x_t = A(\mathcal{L})x_{t-1} + \varepsilon_t$. Define the componential causal strength from $x_i$ to $x_j$ as:

$$F^c_{i\to j} = \frac{\mathrm{var}\left[a_{ji}(\mathcal{L})x_{i,t-1}\right] + \sum_{k=1,k\neq i}^{N}\mathrm{cov}\left[a_{ji}(\mathcal{L})x_{i,t-1}, a_{jk}(\mathcal{L})x_{k,t-1}\right]}{\mathrm{var}\left[x_{jt}\right]}$$

where $\mathrm{var}[\cdot]$ is the variance and $\mathrm{cov}[\cdot,\cdot]$ denotes the covariance of two random variables. Similarly, define the driving force from the noise term as:

$$F^c_{noise_j\to j} = \frac{\mathrm{var}\left[\varepsilon_{jt}\right]}{\mathrm{var}\left[x_{jt}\right]}.$$

Note that in both definitions, the denominator var$[x_{jt}]$ is a normalization factor which can be expressed as:

$$\mathrm{var}\left[x_{jt}\right] = \sum_{k=1}^{N} \mathrm{var}\left[a_{jk}(\mathcal{L})x_{k,t-1}\right] + \sum_{k,l=1,k\neq l}^{N} \mathrm{cov}\left[a_{jk}(\mathcal{L})x_{k,t-1}, a_{jl}(\mathcal{L})x_{l,t-1}\right] + \mathrm{var}\left[\varepsilon_{jt}\right].$$

It is easy to verify that $\sum_{i=1}^{N} F^{c}_{i\to j} + F^{c}_{noise_j\to j} = 1$ for $j = 1, 2, \cdots, N$. The quantity $F^{c}_{i\to j}, i \neq j$ is unbounded and can be both positive and negative. The same applies to $F^{c}_{j\to j}$. $F^{c}_{noise_j\to j}$ is between 0 and 1.

To illustrate this new componential Granger causality approach, consider the following two dimensional autoregressive model:

$$\begin{cases} x_t = ax_{t-1} + 0.9y_{t-1} + \varepsilon_t \\ y_t = 0.9y_{t-1} + \eta_t \end{cases} \quad (9)$$

where $a$ is a parameter, and $\varepsilon_t$ and $\eta_t$ are uncorrelated white noise with zero mean and unit variance. Classical Granger causality considers only the effect of $y$ on $x$, and expresses this relative to the noise. This makes the classical measure very sensitive to the variance of the noise. We show how componential Granger analysis treats also the effect of $x$ on $x$ in this system; the effects of the interaction between $x$ and $y$ on $x$, and measures separately the effect of the noise on $x$. Further, we note that the classical definition of Granger causality imposes no restrictions on the total causal strength from different terms, and show how these sum to 1 with componential Granger causality analysis.

To compare the componential measure with the classical measure of Granger causality, for each value of the parameter $a$, model (9) is simulated to generate 1000 realizations of 1000 time points in order to provide surrogate resampling to construct a confidence interval to assess the significance of each component of componential Granger causality. Fig. 1 shows the means and standard deviations (based on this bootstrap analysis) of the causal strengths for each $a$ obtained from the componential definition (Fig. 1a, b) and the classical definition (Fig. 1c). It can be seen that the causal influences from different sources can be measured with the componential definition. The causal strength from $y$ to $x$ varies with the change of $a$ in the componential model, in contrast to the classical definition. Note that the strongest causal influence from $y$ to $x$ does not occur at $a = 0$ but at some negative value of $a$ which corresponds to a negative causal influence from $x$ to itself. This is due to the fact that some negative values of $a$ lead to an anti-correlation between $x$ and $y$ (see the green line) which counteract each other and reduce the contribution of $x$ to the total causality which is unity. A property of classical Granger causality is that it is unaffected by changing the intrinsic state of the affected system, whereas componential Granger causality can detect this change. We also note that classical Granger causality is effectively a signal to noise ratio approach, whereas componential Granger causality is a systematic approach to measure the effects of cross, self, and noise contributions to causality.

We emphasize in relation to Fig. 1 that classical Granger causality does correctly identify the causal influence from $y$ to $x$, regardless of the values of $a$. Also, Fig. 1b shows that the causal influence from $y$ to $x$ is also identified using componential Granger causality. Moreover, only with componential Granger causality is the causal strength shown to change with the value of $a$. When $a$ takes different values, the intrinsic state of the system changes (i.e. with different values of $a$, the past history of $x$ will have more or less contribution on the current state of itself). This means that componential Granger causality can not only correctly detect the causal influence, but can also provide more information about the system under investigation. That is a reason why we propose componential Granger causality: it takes a more comprehensive and systematic approach to the measurement of Granger causality.

*fMRI investigation*

*Overall design*

We used the identical taste stimulus, 0.1 M monosodium glutamate (MSG) with 0.005 M inosine monophosphate (see de Araujo et al., 2003a), referred to throughout this paper for brevity as monosodium glutamate, in two different types of trials. A trial started 5 s before the taste delivery with the visual attentional instruction either "Remember and Rate Pleasantness" or "Remember and Rate Intensity", which was shown until the end of the taste period. The 0.75 ml taste stimulus was delivered at t = 5 s. The taste period was from t = 5 s until t = 14 s, and in this period a red cross was also present indicating that swallowing should not occur. The differences between the activations in this period were a measure of the top–down selective attention instructions while the taste was being delivered. (We note that in order to utilize top–down attention, one needs to hold the object of attention in mind, in this case pleasantness or intensity. This requires a short-term memory. Short-term memory is thus a sine qua non of selective attention (Rolls, 2008; Rolls and Deco, 2002), and it is the source of this top–down bias from a short-term memory system in which we are interested in this investigation.) After the end of the taste period the visual instruction and red cross were turned off, and a green cross was shown cueing the subject to swallow. After 2 s a tasteless rinse was delivered with a red cross, and the rinse period was from t = 16 until t = 23 s, when the green cross appeared to cue a swallow. After this the rating of pleasantness or intensity was made using button-press operated visual analog rating scales ranging continuously from +2 (very pleasant) to −2 (very unpleasant) for pleasantness, and 4 (intense) to 0 (very weak) for intensity as described previously (Rolls et al., 2003).

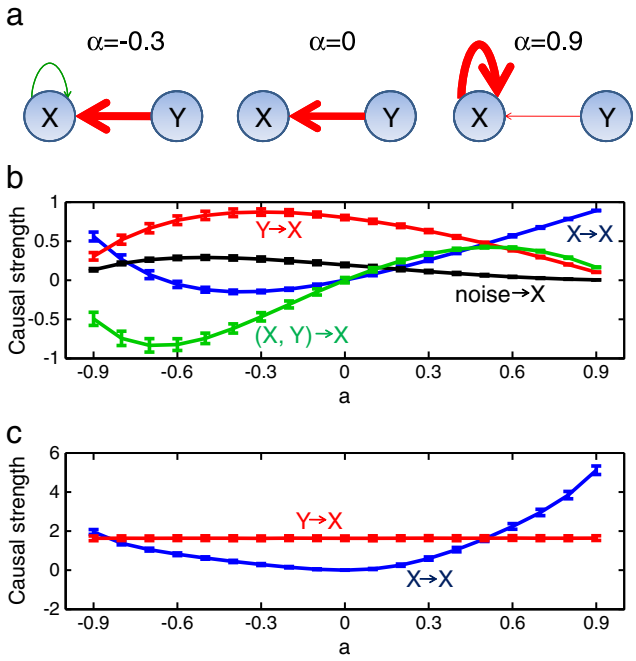


**Fig. 1.** Simulations of the illustrative model defined in Eq. (9) to compare the performance of componential Granger causality (a, b) and classical Granger causality (c). (a) The connectivity pattern of the illustrative model when $a = -0.3$, $a = 0$ and $a = 0.9$. The width of the arrow is proportional to the causal strength measured by componential Granger causality. Red arrows indicate positive causal influences, while the green arrow indicates a negative causal influence. (b) The results of the componential Granger causality approach. $(x, y)\to x$ is Component 3, the interaction term, defined in Eq. (4). The other measures are the componential Granger causality measures defined in Eqs. (6)–(8). (c) The results of the classical Granger causality approach.

These two trial types, in which the instructions were to remember and rate either pleasantness or intensity, were those crucial to the present investigation and its hypotheses. They were interspersed in random permuted sequence with other trials that were part of a different investigation in which there was no pre-trial instruction or attention period, and cognitive modulation by word labels of taste and flavor processing in the brain was being investigated (Grabenhorst et al., 2008). The tastes used on these other trials were monosodium glutamate at concentrations of 0.1 and 0.4 M, the flavors were 0.1 M monosodium glutamate and vegetable odor, and the word labels for the taste condition were "Rich and delicious taste" or "monosodium glutamate". Each of the seven trial types was presented in random permuted sequence 9 times. This general protocol and design has been used successfully in previous studies to investigate taste cortical areas (de Araujo et al., 2003c; Grabenhorst et al., 2008; McCabe and Rolls, 2007; O'Doherty et al., 2001). As 7 trial types were being run in the scanner at the same time, and included different stimuli (Grabenhorst et al., 2008), and no instructions were given about the number of stimuli being used, or that the stimuli were the same on the "Remember and Rate Intensity" and "Remember and Rate Pleasantness" trials, the participants simply had to concentrate on following the instructions about what aspect of the taste stimulus, intensity or pleasantness, had to be rated on that trial.

In a previous analysis of this dataset we focused on identifying the target areas where top–down attentional effects influence responses to taste stimuli (Grabenhorst and Rolls, 2008). Then, with PPI methodology (Friston et al., 1997), we searched for brain areas that might provide the source of the top–down modulation. We identified an anterior lateral prefrontal cortex region at Y = 53 mm in which the correlation with activity in the orbitofrontal cortex seed region was greater when attention was to pleasantness than to intensity (Grabenhorst and Rolls, 2010). Conversely, in a more posterior region of lateral prefrontal cortex at Y = 34 the correlation with activity in the anterior insula seed region was greater when attention was to intensity than to pleasantness. Here we investigate the directionality of the influences between these areas using the componential Granger causality analysis, to obtain evidence on whether the prefrontal cortex is having top–down effects on the orbitofrontal cortex and on the insula.

*Participants*

Twelve healthy volunteers (6 males and 6 females, age range 21–35) participated in the study. Ethical approval (Central Oxford Research Ethics Committee) and written informed consent from all subjects were obtained before the experiment. The subjects had not eaten for 3 h before the investigation. With our sample size of 12 participants, statistically significant effects in random effects group analyses were found (Grabenhorst and Rolls, 2008, 2010).

*Stimuli and stimulus delivery*

The taste stimulus was monosodium glutamate (0.1 M MSG and 0.005 M inosine monophosphate). We included a tasteless control solution containing the main ionic components of saliva (25 mM KCl + 2.5 mM $NaHCO_3$) which when subtracted from the effects produced by the taste stimulus allowed somatosensory and any mouth movement effects to be distinguished from the effects purely related to taste (de Araujo et al., 2003a; O'Doherty et al., 2001). This is an important control condition that we have pioneered to allow taste areas to be shown independently of any somatosensory effects produced by introducing a fluid into the mouth (de Araujo et al., 2003a, 2003b; O'Doherty et al., 2001). For the PPI analyses described in this paper, explicit subtraction of effects related to the rinse was not applied, as the rinse effect was common to each of the two attentional conditions involved in the analyses.

The stimuli were delivered to the subject's mouth through four teflon tubes (one for each of the 3 taste or flavor stimuli, and a separate tube for the tasteless rinse control) that were held between the lips. Each teflon tube of approximately 3 m in length was connected to a separate reservoir via a syringe and a one-way syringe activated check valve (Model 14044-5, World Precision Instruments, Inc), which allowed 0.75 ml of any stimulus to be delivered at the time indicated by the computer.

*fMRI data acquisition*

Images were acquired with a 3.0-T VARIAN/SIEMENS whole-body scanner at the Centre for Functional Magnetic Resonance Imaging at Oxford (FMRIB), where 27 T2* weighted EPI coronal slices with in-plane resolution of 3 × 3 mm and between plane spacing of 4 mm were acquired every 2 s (TR = 2). We used the techniques that we have developed over a number of years (de Araujo et al., 2003a; O'Doherty et al., 2001) and as described in detail by Wilson et al. (2002) we carefully selected the imaging parameters in order to minimize susceptibility and distortion artifact in the orbitofrontal cortex. The relevant factors include imaging in the coronal plane, minimizing voxel size in the plane of the imaging, as high a gradient switching frequency as possible (960 Hz), a short echo time of 28 ms, and local shimming for the inferior frontal area. The matrix size was 64 × 64 and the field of view was 192 × 192 mm. Continuous coverage was obtained from + 62 (A/P) to − 46 (A/P). A whole brain T2* weighted EPI volume of the above dimensions, and an anatomical T1 volume with coronal plane slice thickness 3 mm and in-plane resolution of 1 × 1 mm was also acquired.

*fMRI data analysis*

The imaging data were analyzed using SPM5 (Statistical Parametric Mapping, Wellcome Trust Centre for Neuroimaging, London). Preprocessing of the data used SPM5 realignment, reslicing with sinc interpolation, normalization to the MNI coordinate system (Montreal Neurological Institute) (Collins et al., 1994), and spatial smoothing with a 6 mm full width at half maximum isotropic Gaussian kernel. Time series non-sphericity at each voxel was estimated and corrected for (Friston et al., 2002), and a high-pass filter with a cut-off period of 128 s was applied. To investigate task dependent correlations between brain areas during the taste period a Finite Impulse Response (FIR) analysis was performed as implemented in SPM, in order to make no assumption about the time course based on the temporal filtering property of the hemodynamic response function (Gottfried et al., 2006; Henson, 2003; Yacubian et al., 2006). We used 16 delta functions in the FIR analysis, spaced at intervals of the TR (2 s) and starting 2 s before the onset of the visual cue at time − 5 s instructing the subject to "Remember and rate pleasantness" (or intensity) on that trial. This basis set considers each time bin after stimulus onset individually to model the BOLD (blood oxygen level dependent) response and can capture any possible shape of response function up to a given frequency limit. In this model, the parameter estimate for each time bin represents the average BOLD response at that time. These parameter estimates are directly proportional to the BOLD signal. Data were analyzed statistically for each subject individually (first-level analysis) and for the group (second level analysis). At the single-subject level, the parameter estimate maps for the FIR analysis for the time delays at which taste-related activations were expected were estimated (in our case, the FIR values for times starting 6 s and 8 s after taste delivery were used, and are consistent with the hemodynamic delays inherent in the BOLD response (Henson, 2003)). Following smoothness estimation (Kiebel et al., 1999), in the first stage of analysis condition-specific experimental effects (parameter estimates, or regression coefficients, pertaining to the effect size of the PPI) were obtained via the general linear model (GLM) in a voxel-wise manner for each subject. The statistical parametric maps from each individual dataset were then entered into second-level, random effects analyses accounting for both scan-to-scan and subject-to-subject variability. More precisely, the sets of individual statistical maps corresponding to a specific effect of interest were entered

as covariates in multiple regression models as implemented in SPM5, and the corresponding group effects were assessed by applying linear contrasts (again following smoothness estimation) to the (second-level) parameter estimates generating a t-statistics map for each group effect of interest. The SPM contrast analyses performed tested for a difference between paying attention to pleasantness vs intensity and vice versa. For the contrast analysis reported in Grabenhorst and Rolls (2008) the main regressors of interest in our GLM modeled the onset of the taste stimuli on the attention to pleasantness and attention to intensity trials for the FIR times corresponding to the taste period as described above starting at 6 and 8 s using t-contrasts as implemented in SPM. This type of approach was chosen because the 6 and 8 second bins were identified (by examining average FIR BOLD response time courses in the present dataset as well as in other datasets where taste-related responses were measured) as reflecting the peak of the taste-related BOLD response in the selective attention design in taste-related areas of interest including the insular taste cortex and orbitofrontal cortex without also reflecting any effects related to the attention instructions at the start of the trial. Other task periods, including specifically the attentional instruction period, swallowing periods, rinse period, and rating periods, as well as other trial types not related to the present investigation were modeled as effects of no interest. The group-level t-tests were performed with 11 degrees of freedom.

We performed PPI analyses (Friston et al., 1997; Gitelman et al., 2003) to investigate how activity in pairs of brain regions is modulated as a function of task (attentional condition). The methods used and the results obtained with this dataset have been described previously (Grabenhorst and Rolls, 2010). The a priori defined areas of interest for which we report results included brain areas where activations to taste stimuli have been found in previous studies including the medial and lateral orbitofrontal cortex, the pregenual part of the cingulate cortex, and the taste and oral somatosensory parts of the insular cortex (de Araujo et al., 2003a, 2003c; Grabenhorst et al., 2008; McCabe and Rolls, 2007; Nitschke et al., 2006; O'Doherty et al., 2001; Schoenfeld et al., 2004); and areas of the lateral prefrontal cortex where activations related to task set, attentional instructions, and remembering rules that guide task performance have been found, including specifically parts of the middle and inferior frontal gyrus (Beck and Kastner, 2009; Bengtsson et al., 2009; Deco and Rolls, 2005a; Kouneiher et al., 2009; Rossi et al., 2009; Sakai and Passingham, 2003, 2006; Veldhuizen et al., 2007).

We applied small volume (false discovery rate) corrections for multiple comparisons for which $p<0.05$ (though the exact corrected probability values are provided) (Genovese et al., 2002) with a radius corresponding to the full width at half maximum of the spatial smoothing filter used. Peaks are reported for which $p<0.05$ corrected for false discovery rate, though the actual corrected probability values are given in the text. In addition to the statistical criterion just described for a significant effect calculated for the peak voxel of a region of activation in an a priori defined region based on earlier findings, we used the additional statistical test (see Gottfried et al., 2002; O'Doherty et al., 2003, 2006) that the results reported were significant in global contrast and/or correlation analyses using the criterion of $p<0.001$ uncorrected for multiple comparisons, and these additional statistics confirmed the same effects in the a priori regions in all cases in this paper unless otherwise stated.

For the componential Granger causality analyses described here, no temporal smoothing was used, and the raw activation values were extracted from the normalized and realigned volumes (the wr* files in SPM). Voxels were selected for based on statistically significant results for a contrast or correlation in the conventional SPM analyses, the results of which are reported elsewhere (Grabenhorst and Rolls, 2008, 2010). The mean BOLD signals in 33 voxels each $3\times3\times3$ mm within a sphere of radius 2 voxels were used in the causality analyses.

To measure the directionality of the causal effects of interest, we performed statistical comparisons of the difference between the componential Granger causality measure in Eq. (6) for the direction from the anterior prefrontal cortex to the orbitofrontal cortex and in the opposite direction. We did the same for effects from the posterior prefrontal cortex to and from the taste insula. Although such differences of Granger causality measures may be approximately Gaussian distributed (Schippers and Keysers, 2011), we chose to use non-parametric tests, as they make no assumptions about the distribution, and provide exact values. In detail, suppose the top–down and bottom–up effects are the same. Under this null hypothesis, the direction of the causal strength is arbitrary for each subject. Therefore, we can permute the top–down and bottom–up measures for each subject and thus obtain the null distribution of the mean effect by averaging across the subjects. We performed all the possible $2^{12}=4096$ relabelings and constructed the null distribution of the statistic. The p value is given by counting the proportion of the permutation distribution as extreme or more extreme than our observed statistic.

To test whether the unidirectional causality measure is significantly larger than zero, we randomly rearranged the labels of data points and calculated the test statistic. This gives the distribution of the causal measure under the null hypothesis, i.e., that no causal influence exists. Bonferroni correction was used to control for multiple comparisons. The significance level used after the Bonferroni correction was $p<0.05$. The results shown in the Tables and reflected in the Figures are based on these non-parametric tests.

## Results

As we have reported previously (Grabenhorst and Rolls, 2008), a contrast of trials where attention was being paid to taste pleasantness with trials where attention was to intensity revealed significant effects in the orbitofrontal cortex [−6 14 −20] and a region to which it projects, the pregenual cingulate cortex [−4 46 −8]. The reverse contrast of trials where attention was to intensity vs trials where attention was to pleasantness revealed significant effects in the right anterior insular taste cortex [42 18 −14] and the right mid insular cortex [40 −2 4]. To investigate task-dependent functional connectivity of these areas with other brain regions we performed PPI analyses using the above brain areas as seed regions (see Fig. 2b). We found using the PPI analyses that a region of the anterior lateral prefrontal cortex (antLPFC) at $Y=54$ has correlations with the orbitofrontal cortex taste area that were greater during attention to the pleasantness of taste than to the intensity of taste; that a region of the posterior lateral prefrontal cortex (postLPFC) at $Y=34$ has correlations with the anterior insular (primary) taste cortex that were greater during attention to the intensity of the taste than to the pleasantness of the taste; and that a region of the lateral prefrontal cortex at $Y=34$ has correlations with the mid insular area of cortex activated by taste that were greater during attention to the intensity of the taste than to the pleasantness of the taste (see Fig. 3) (Grabenhorst and Rolls, 2010). These effects are shown in Table 1, and the brain areas involved, which were those studied in the present investigation, are indicated in Figs. 2 and 3.

In the present investigation, to assess the directionality of the effects between these prefrontal cortex areas, and the orbitofrontal cortex and insular cortical areas, we extracted the time series of the BOLD signal from 33 voxels centered at the coordinates shown in Table 1. These time series had no temporal smoothing, but did have 6 mm spatial smoothing applied as standard to this dataset (Grabenhorst and Rolls, 2010). (We checked that temporal aliasing did not influence the findings by performing a control analysis with a 3-point temporal smoothing filter applied, and this showed similar, though smaller, effects to those described here.) Each time series consisted of 9 trials each with 18 BOLD signal data points each 2 s apart (the TR) starting on each trial at the onset of the instruction to pay attention to the pleasantness or to the intensity of the taste. (The taste stimulus, MSG, was identical on all trials. For the whole of the analysis period there was a top–down attentional effect, and interactions between the two brain areas are being compared.

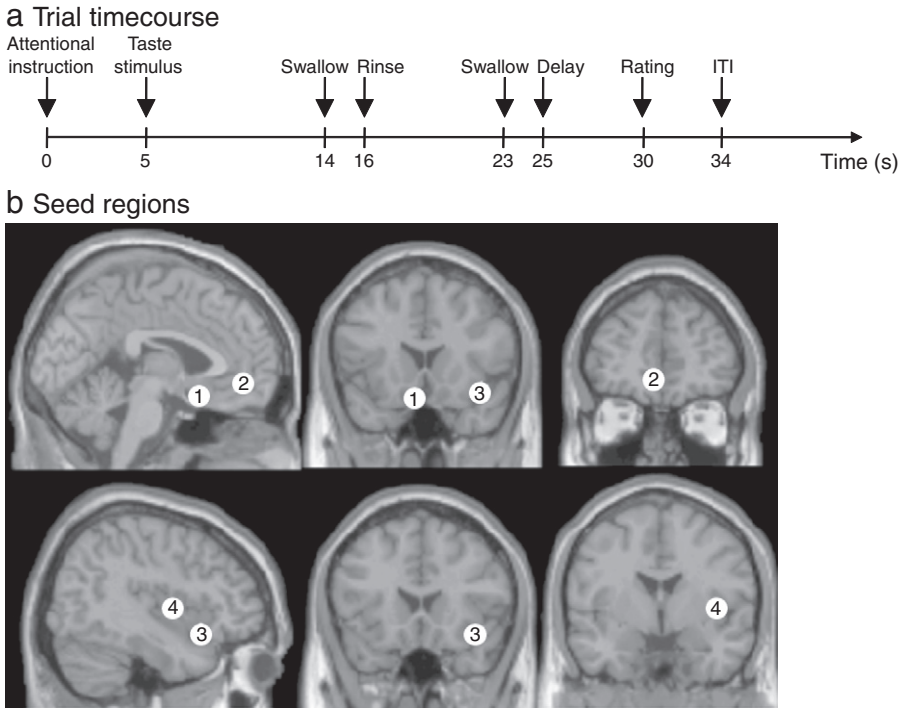


**Fig. 2.** a. The time course of a trial. The taste stimulus and the rinse were delivered at the times shown. ITI — start of the inter-trial interval. b. The seed regions for the PPI analyses. The orbitofrontal cortex (1) was the seed region at which attention to the pleasantness of a taste modulated the response to a taste. The anterior insula (3) and mid insula (4) were seed regions at which attention to the intensity of a taste modulated the response to a taste. The pregenual cingulate cortex (2) was not included in this investigation.

Thus we do not try to separate out different periods of a trial in which the taste was present or not (cf. Bressler et al., 2008). The Granger causality analysis performed would reflect for example stronger effects in one direction than another over this timecourse, and we then analyzed whether these effects depended on whether selective attention was being paid to affective value or intensity.) For each participant, one value is obtained for the Granger coefficient for each direction between a pair of brain areas. The statistics then test for significant differences across subjects in the forward and backward directions between a pair of brain areas. The results of the componential Granger analysis are shown in Tables 2 and 3. Table 2 shows $F^{c}_{y\rightarrow x}$. Table 3 shows the values of the Components that contribute to this, which are Component 1, and half of Component 3.

First, it can be seen from Table 2 that there is a significantly greater effect from the anterior lateral prefrontal cortex to the orbitofrontal cortex (OFC) than vice versa, and that this effect is present during attention to pleasantness (bold font in the top row in Table 2), but not during attention to intensity. The interpretation is that the anterior lateral prefrontal cortex influences the OFC, but only during attention to pleasantness. The implication is that the anterior lateral prefrontal cortex is the brain site of the top–down attentional influence on the

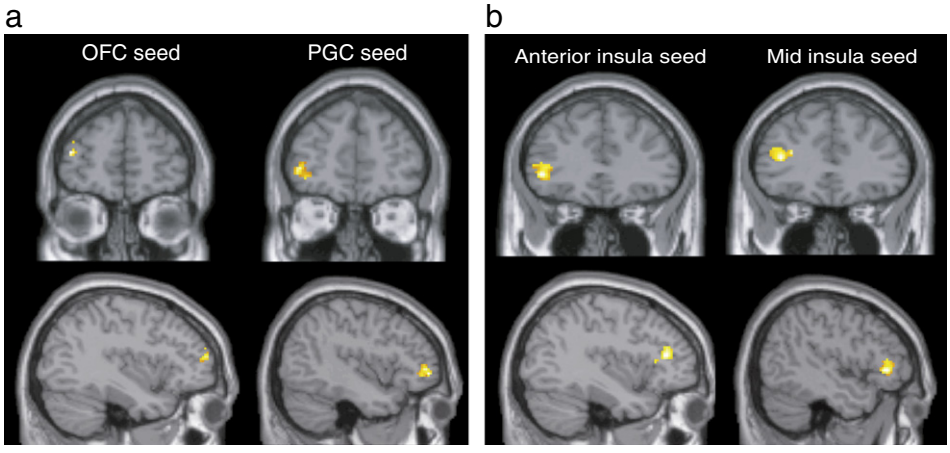


**Fig. 3.** a. Regions of the anterior lateral prefrontal cortex at which the correlation with activity in the orbitofrontal cortex and pregenual cingulate cortex seed regions was greater when attention was to pleasantness compared to when attention was to intensity. Left. Anterior lateral prefrontal cortex region at Y = 54 [−40 54 14] at which the correlation with activity in the orbitofrontal cortex (OFC) seed region was greater when attention was to pleasantness compared to when attention was to intensity. Right. Anterior lateral prefrontal cortex region at Y = 50 [−42 50 −2] at which the correlation with activity in the pregenual cingulate cortex seed region was greater when attention was to pleasantness compared to when attention was to intensity. b. Regions of lateral prefrontal cortex at which the correlation with activity in insular cortex seed regions was greater when attention was to intensity compared to when attention was to pleasantness. Left. Lateral prefrontal cortex region at Y = 34 [−38 34 14] at which the correlation with activity in the anterior insula seed region was greater when attention was to intensity compared to when attention was to pleasantness. Right. Lateral prefrontal cortex region at Y = 34 [−46 34 0] at which the correlation with activity in the mid insula seed region was greater when attention was to intensity compared to when attention was to pleasantness.

**Table 1**
Brain regions where significant effects in psychophysiological interaction analyses were found. All effects are significant at $p<0.05$ (small volume correction) and $p<0.001$ (uncorrected in whole brain analyses) except where otherwise stated. * $p<0.05$ (small volume correction) and $p=0.002$ (uncorrected in whole brain analysis).

| Brain region | X | Y | Z | z-value | p-value |
|---|---|---|---|---|---|
| *Orbitofrontal cortex seed region [−6 14 −20] for attention to pleasantness vs attention to intensity* | | | | | |
| Ant lateral prefrontal cortex | −40 | 54 | 14 | 3.32 | 0.029 |
| *Anterior insular cortex seed region [42 18 −14] for attention to intensity vs attention to pleasantness* | | | | | |
| Posterior lateral prefrontal cortex | −38 | 34 | 14 | 2.87 | 0.049* |
| *Mid insular cortex seed region [40 −2 4] for attention to intensity vs attention to pleasantness* | | | | | |
| Lateral prefrontal cortex | −46 | 34 | 0 | 3.35 | 0.008 |

OFC during attention to pleasantness, when activations to the taste stimulus are higher when attention is to pleasantness than to intensity. It is also shown (in the second row of Table 2) that the posterior lateral prefrontal cortex area ($y=34$, identified because it was seeded from the anterior insula, the primary taste cortex) also had a significant top–down effect on the orbitofrontal cortex during attention to pleasantness but not to intensity. The implication is that the anterior lateral prefrontal area at $y=54$ is not entirely functionally independent of the posterior lateral prefrontal cortex area at $y=34$ in its effects during pleasantness on the OFC. (In the PPI analyses, the anterior and posterior LPFC sites had consistently different peaks that were seeded from the orbitofrontal cortex and anterior insula taste cortex respectively (Grabenhorst and Rolls, 2010).)

Second, it can be seen from Table 2 that the componential Granger coefficient is significantly larger from the posterior lateral prefrontal cortex to the anterior insula (primary taste cortex) than vice versa, and that this effect is found only when attention is paid to intensity and not when it is paid to pleasantness. The interpretation is that the posterior lateral prefrontal cortex ($y=34$) influences the anterior insular taste cortex, but only during attention to intensity and not during attention to pleasantness. The implication is that the posterior lateral prefrontal cortex is the brain site of the top–down attentional influence on the anterior insula primary taste cortex during attention to intensity. The finding shown in Table 2 that causal effects are not detected from the anterior lateral prefrontal cortex to or from the anterior insular (primary taste) cortex (during selective attention to either pleasantness or intensity) indicates that there is some independence during attention to intensity of the functions of the anterior lateral ($y=54$) and posterior lateral ($y=34$) prefrontal cortical areas in their top–down effects that control attention to the intensity of a taste stimulus.

This dissociation of top–down control influencing the OFC more strongly in the pleasantness than the intensity condition, and top–down control influencing the anterior taste insula more strongly in the intensity than the pleasantness condition was tested in a two-factor ANOVA. One factor was pleasantness vs intensity. The second factor was net causality for the AntLPFC to OFC effects vs the net causality for the PostLPFC to AntIns effects. (Net causality is the difference in the cross componential causality of Eq. (6) between the two directions.) This hypothesized interaction was significant ($p<0.04$), consistent with the hypothesis that the AntLPFC effect on the OFC is relatively greater during attention to pleasantness, and that the PostLPFC effect on the AntIns is relatively greater during attention to intensity. These results are illustrated in Fig. 4, in which the net causality is positive when it reflects a greater top–down effect from the anterior lateral prefrontal cortex to the orbitofrontal cortex; and from the posterior lateral prefrontal cortex to the anterior insula.

Third, it can be seen from Table 2 that the lateral prefrontal cortex area ($y=34$) seeded from the mid insular taste region has a stronger effect in the direction from the LPFC to mid-insula, and only during attention to intensity and not during attention to pleasantness, though this did not quite reach significance across the 12 participants ($p=0.054$). This is consistent with this LPFC area exerting a top–down attentional modulation on the mid-insular cortex region activated by taste during selective attention to intensity but not to pleasantness.

The effects shown by the componential Granger causality measure $F^c_{y\to x}$ shown in Table 2 are now further interpreted in terms of the values shown in Table 3 of the four components (Eqs. (2)–(5)). With respect to the anterior lateral prefrontal cortex/orbitofrontal cortex, when attention is being paid to pleasantness, both the cross Component (1), and the interaction Component (3), are significantly higher for the direction from the anterior lateral prefrontal cortex to the orbitofrontal cortex. The interpretation is that there is a top–down influence of the anterior lateral prefrontal cortex on the orbitofrontal cortex, and that this depends on whether the orbitofrontal cortex is activated (as shown by the interaction term, i.e. Component 3). This interaction term is not measured by classical Granger causality analysis, and it is an important aim and advantage of componential Granger causality that it can measure such interaction effects. It is an interesting effect neurophysiologically, and probably corresponds to a situation in which the top down effects only become evident when this processing stream (orbitofrontal cortex and anterior lateral prefrontal cortex) is active. This interaction effect is not found in this affective (orbitofrontal cortex) stream when attention is being paid to intensity.

A significant interaction term (Component 3) was also found for the effect of the posterior LPFC on the anterior insula during attention to intensity. The interpretation is that the top–down attentional effect of the posterior LPFC on the anterior insula (primary taste cortex) depends on whether the anterior insula is activated. There is also a significant interaction term in the same pathway during attention to pleasantness. As shown in Table 3, the interaction term, Component 3, can be negative. If it is (significantly) negative, the interpretation

**Table 2**
Brain regions tested for top–down vs. bottom–up effects with the componential Granger causality analysis. The magnitude of the coefficient $F^c_{y\to x}$shown indicates the strength of the directionality measure averaged across the 12 participants, with the p value obtained by permutation resampling shown to the right of the two directionality coefficients being compared. The values in the Table show where an effect was predicted in the direction of lateral prefrontal cortex to OFC or Insula. The values in brackets are the p values of a test for whether the coefficient is significantly larger than zero. Values in bold font in the brackets show where these are significant after Bonferroni correction. The values in the columns headed 'p value' are for a permutation test between the values in the preceding 2 columns for 'top–down' vs 'bottom–up' effects, with significant differences shown in bold font.

| Orbitofrontal cortex seed region [−6 14 −20] | Attention to pleasantness | | | Attention to intensity | | |
|---|---|---|---|---|---|---|
| | To OFC | From OFC | p value | To OFC | From OFC | p value |
| Anterior lateral prefrontal cortex [−40 54 14] | **0.0252 (0.0001)** | 0.0060 (0.5646) | **0.0005** | **0.0164 (0.0023)** | 0.0150 (0.0065) | 0.3962 (NS) |
| Posterior lateral prefrontal cortex [−38 34 14] | **0.0278 (0.0001)** | 0.0069 (0.4141) | **0.0024** | 0.0134 (0.0163) | 0.0085 (0.2171) | 0.3049 (NS) |
| Anterior insular cortex seed region [42 18 −14] | To AntINS | From AntINS | p value | To AntINS | From AntINS | p value |
| Anterior lateral prefrontal cortex [−40 54 14] | 0.0106 (0.0818) | **0.0198 (0.0002)** | 0.1619 (NS) | **0.0180 (0.0010)** | 0.0103 (0.0857) | 0.1892 (NS) |
| Posterior lateral prefrontal cortex [−38 34 14] | **0.0247 (0.0001)** | 0.0124 (0.0363) | 0.1184 (NS) | **0.0214 (0.0001)** | 0.0078 (0.3032) | **0.0356** |

**Table 3**
The values of the four components of componential Granger causality as defined in Eqs. (2)–(5) for this fMRI investigation. The cross and interaction terms within the Table shows the values directed to that brain region in the Table from the other brain region in the same pair of rows of the Table. The p values shown are for whether there is a significant difference between the values of the components in the two rows immediately above in the same column, with significant differences shown in bold font.

| | Attention to pleasantness | | | | Attention to intensity | | | |
|---|---|---|---|---|---|---|---|---|
| | Self | Cross | Interaction | Noise | Self | Cross | Interaction | Noise |
| | Comp 2 | Comp 1 | Comp 3 | Comp 4 | Comp 2 | Comp 1 | Comp 3 | Comp 4 |
| Orbitofrontal cortex seed region [−6 14 −20] | 0.0590 | 0.0196 | 0.0113 | 0.9102 | 0.0684 | 0.0136 | 0.0056 | 0.9123 |
| Anterior lateral prefrontal cortex [−40 54 14] | 0.1795 | 0.0053 | 0.0013 | 0.8139 | 0.2176 | 0.0117 | 0.0066 | 0.7641 |
| p value | | **0.0007** | **0.0227** | | | 0.2976 (NS) | 0.4497 (NS) | |
| Anterior insular cortex seed region [42 18 −14] | 0.1715 | 0.0154 | 0.0184 | 0.7946 | 0.1727 | 0.0153 | 0.0121 | 0.7999 |
| Posterior lateral prefrontal cortex [−38 34 14] | 0.3000 | 0.0175 | −0.0103 | 0.6928 | 0.3204 | 0.0088 | −0.0020 | 0.6728 |
| p value | | 0.4182 (NS) | **0.0078** | | | 0.1445 (NS) | 0.0674 (NS) | |

is that variation in $y$ is out of phase with $x$ (or the opposite of $x$). The effect of this is that variation in $y$ would decrease the variation in $x$. The interaction term is thus important in interpreting whether a change in $y$ in one direction produces a change in the same direction (interaction positive) or the opposite direction (interaction negative). If oscillations were present and in phase, the cross term would be greater than zero if one affects the other (or zero if it does not), and the interaction term, Component 3, could be positive, or could be zero. If the oscillations were out of phase, and $y$ affects $x$, the cross component would be greater than zero, and the interaction term typically would be negative. (We note that an interaction term, Component 3, will only be non-zero if both the cross Component (1) and the auto Component (2) are larger than zero.)

With respect to the anterior (taste) insula/posterior lateral prefrontal cortex, it is shown in Table 3 that when attention is being paid to intensity, the interaction term is larger (though not quite significant, $p = 0.06$) for the direction from the posterior lateral prefrontal cortex to the anterior insula. This is consistent with the hypothesis that there is a top down effect of the posterior lateral prefrontal cortex on the taste cortex in the anterior insula only when the insular part of that processing stream for the intensity of taste stimuli is being activated. Interestingly, an interaction effect is also found in the same direction during attention to pleasantness. This implies that the taste insula is being affected by top–down signals during attention to pleasantness.

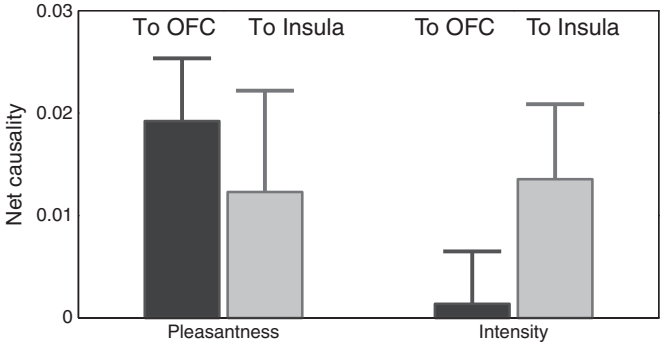


**Fig. 4.** The net causality for the effects from the Anterior lateral prefrontal cortex to the orbitofrontal cortex (OFC); and from the posterior lateral prefrontal cortex to the anterior insula (primary taste cortex) when attention was to the pleasantness of the taste or the intensity of the taste. Net causality is the difference in the cross componential causality of Eq. (6) between the two directions. Means and standard errors are shown, and with the within-subjects design the interaction was significant ($p < 0.04$). The net causality shown is positive when it reflects a greater top–down effect from the anterior lateral prefrontal cortex to the orbitofrontal cortex; and from the posterior lateral prefrontal cortex to the anterior insula.

In Table 3, the noise Component (4) measures what in terms of the variations in $x$ cannot be accounted for by prior changes in $y$ and $x$. It is in a sense the unattributed variance in $x$. This draws attention to the fact that the magnitude of the causal influence from y to x is measured on a useful scale of absolute value which for the three terms in componential Granger causality (Eqs. (6)–(8)) have been defined to sum to 1. Thus the amount of variation in the system that can be accounted for by changes in $x$ and $y$ can be measured on a scale that can be compared between different systems, as the measure is always in the range 0 to 1. A large value for the cross term, Component 1, of for example 0.1, would indicate a strong effect of $y$ on $x$. We note that in the system analyzed here, fMRI data, the absolute values of Components 1–3 are not large; this is related to the fact that fMRI signals are noisy, and vary from trial to trial.

In Table 3, the self or auto Component (2) is generally quite large, typically between 0.05 and 0.3. The interpretation is that the previous history of the signal in a brain area is a good predictor of the future signal, which in the context of this fMRI investigation is reasonable.

We can think of the signal to noise ratio in two ways with componential Granger causality. First, the ratio of $F^c_{y\to x} + F^c_{x\to x}$ to $F^c_{noise_x\to x}$ can be thought of as a signal to noise ratio in which the measured effects of $y$ on $x$ and $x$ on $x$ are considered as the signal, and the otherwise unattributed effects as noise. Second, the ratio of $F^c_{y\to x}$ to $F^c_{x\to x} + F^c_{noise_x\to x}$ might possibly be thought of as a signal to noise ratio in which the effect of the signal input $y$ on the $x$ system is considered, with the noise and the effects of $x$ on $x$ being what is not contained in the signal.

Fig. 5 summarizes some of the effects found with componential Granger causality analysis. It is based on the analyses shown in Table 2. Values that are significantly larger than zero are shown in Table 2 in bold font, and form the basis of the arrows indicated in Fig. 5. First, we make the point that causality can operate in both directions. For example, there are significant effects in both directions between the orbitofrontal cortex and the anterior LPFC during attention to intensity. Values significantly larger than 0 are indicated by the blue arrows in Fig. 5b. The anatomical basis for this is that in addition to forward connections from the orbitofrontal cortex to the anterior LPFC, there are reciprocal backprojections. Of course whether the connections in either or both directions are engaged is a functional issue, and is what is revealed by componential Granger causality. In some cases the effects in one direction are stronger than in the other. For example, in the same pathway, in Fig. 5a when attention is being paid to pleasantness, the top–down causality is stronger from the anterior LPFC to the orbitofrontal cortex. Red arrows in Fig. 5 indicate stronger causality in one direction than the other. Arrows are only shown in Fig. 5 when the t tests remain significant with a Bonferroni correction (which in this case was for 8 comparisons). To help with the interpretation of these causality values, values of the four components that are significantly different from zero are indicated in Table 3.

The results on the same dataset for classical Granger causality analysis are shown in Table 4. The classical Granger analysis values

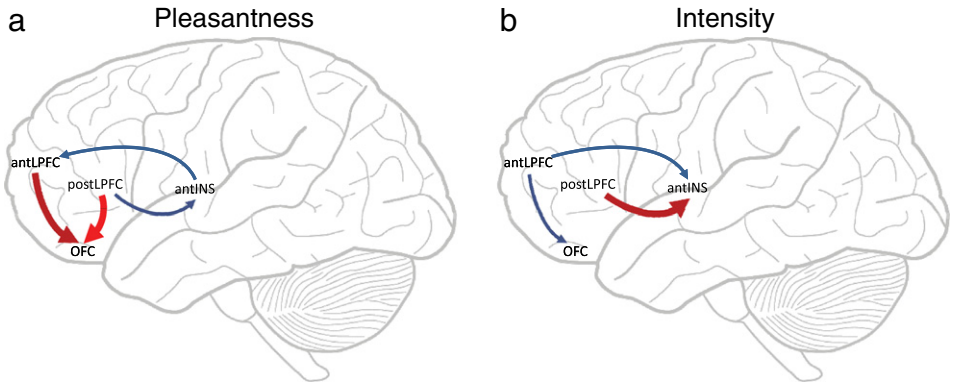


**Fig. 5.** Schematic diagram of the causal influences from componential Granger causality analysis shown in this investigation. Significant causal influences from t tests with a Bonferroni correction are marked by blue arrows (i.e. cross-componential Granger causality is greater than 0). Red arrows indicate significant top–down effects exist in addition to significant causal influences (i.e. a significant cross-componential Granger causality that is different in the two directions). a) During attention to pleasantness. b) During attention to intensity. The areas are anterior and posterior lateral prefrontal cortex (antLPFC, postLPFC); orbitofrontal cortex (OFC); and anterior insula cortex (antINS).

in Table 4 are somewhat less significant than for the componential Granger causality analysis in the top part of the Table for the OFC connectivity; and are not significant at all in the bottom part of Table 4 for the classical Granger analysis of the insular taste cortex connectivity. Componential Granger causality may thus we suggest produce in at least some circumstances more significant results than classical Granger causality analysis, with a particular advantage being that it can take into account interaction effects. The absolute magnitude of the causality effects measured were not very large (with either the classical or componential Granger causality approach), and we attribute that to the noisy properties of fMRI signals, and to the fact that top–down attentional influences must not be very strong, for they must only bias, and not dominate, the bottom–up signal (Deco and Rolls, 2005b; Renart et al., 1999; Rolls and Deco, 2002). It must be noted that differences in hemodynamic response functions between brain areas can in principle influence the interpretation of causality measures (Deshpande et al., 2010b). Some studies (e.g. Schippers et al., 2011) have shown by intensive simulations that Granger causality is relatively robust in group-level analysis. In any case, such effects were minimized in the present experimental design in which causal influences were measured under two different attentional conditions, and differences in causal influences between the same pairs of brain areas depended on the attentional condition. However, it will be of interest to investigate how componential Granger analysis performs on neuronal data recorded simultaneously in different brain areas.

We also measured the correlations between the brain areas shown in Table 2, and while these reflected the PPI analyses, were clearly very different from what was shown by the Granger causality analyses. For example, the posterior lateral prefrontal cortex and the orbitofrontal cortex had high correlations (0.10), but the causal influences between them were not significant when Bonferroni corrected, as shown in Table 3. A possible interpretation in such a case is that there is common input to both the regions.

**Table 4**
Brain regions tested for top–down vs. bottom–up effects with the classical Granger analysis. The magnitude of the coefficient indicates the strength of the directionality measure averaged across the 12 participants, with the p value obtained by permutation resampling shown to the right of the two directionality coefficients being compared.

| Orbitofrontal cortex seed region [−6 14 −20] | | Attention to pleasantness | | | Attention to intensity | | |
|---|---|---|---|---|---|---|---|
| | | To OFC | From OFC | p value | To OFC | From OFC | p value |
| Anterior prefrontal [−40 54 14] | lateral cortex | 0.0219 | 0.0066 | **0.0007** | 0.0150 | 0.0161 | 0.4089 (NS) |
| Posterior prefrontal [−38 34 14] | lateral cortex | 0.0253 | 0.0066 | **0.0027** | 0.0109 | 0.0133 | 0.3818 (NS) |
| Anterior insular cortex seed region [42 18 −14] | | To AntINS | From AntINS | p value | To AntINS | From AntINS | p value |
| Anterior prefrontal [−40 54 14] | lateral cortex | 0.0131 | 0.0223 | 0.1145 (NS) | 0.0184 | 0.0104 | 0.1907 (NS) |
| Posterior prefrontal [−38 34 14] | lateral cortex | 0.0190 | 0.0223 | 0.3896 (NS) | 0.0177 | 0.0132 | 0.3032 (NS) |

## Discussion

The results described here with the componential Granger causality analyses show that there is a top–down effect related to selective attention from the lateral prefrontal cortex to the orbitofrontal cortex and insular taste cortex. The direction of the effect could not be established with the previous PPI analyses (Grabenhorst and Rolls, 2010), which rely on correlation, and which cannot detect the direction of effects, and cannot address causality (Friston et al., 1997; Gitelman et al., 2003; Grabenhorst and Rolls, 2010). Moreover, the results show that the top–down influences are selectively switched on: for example from the anterior lateral prefrontal cortex to the OFC during selective attention to the pleasantness of the taste stimuli; and from the posterior lateral prefrontal cortex to the insula during selective attention to the intensity of the taste (see Table 2). Moreover, the componential Granger analyses were in general consistent with the PPI analyses, with for example the source of the prefrontal cortex top–down influence on the taste insula that is directionally selective being from the posterior lateral prefrontal cortex (y = 34) but not from the anterior lateral prefrontal cortex (y = 54) (see Table 2).

Componential Granger analysis is very useful in providing evidence on the directionality of the functional connectivity, including task dependency, whereas diffusion tensor tractography (Klein et al., 2010) informs only about structural connectivity, and provides no evidence on the direction of the connectivity. It is also potentially useful for estimating the relative strengths of forward and backward connections between cortical areas, which are thought to be stronger in the forward than backward direction because the back projections end far out on the apical dendrites of cortical pyramidal cells in layer 1 (Rolls, 2008). Further, it is predicted that during memory recall the back projections (from for example the hippocampus to the neocortex) would become functionally more effective (as there is less forward input to shunt the back projected inputs to neurons (Rolls,

2008)), and this could be investigated with componential Granger causality analysis. Although we have utilized componential Granger causal analysis in a powerful way in this paper by focusing on areas identified in a PPI analysis to investigate the strength of the causal influence under different task (selective attentional) conditions, componential Granger causal analysis could also be useful for simply helping to identify the direction of a functional influence without varying a task.

It would be impractical to apply the method we describe to every voxel in an fMRI study, as the dimensionality would be too large for significant effects. In this study, we therefore took advantage of correlation analyses which were corrected statistically to identify in a PPI analysis pairs of brain regions that were connected (Grabenhorst and Rolls, 2010), and then used componential Granger causality analysis to provide evidence on the directionality and the magnitude of the influence of one area on another. A point that must be borne in mind in any such analysis is that common inputs to two different areas might lead to correlations between the activity in one area and the other, and it is possible, if the time lags to one area were longer or the dynamics slower, that a Granger analysis might suggest a causal influence of one of the two areas on the other, when in fact common inputs but with dynamic differences account for the Granger coefficients.

Componential Granger causality has several advantages over the classical definition. The most significant improvement is to assess the causal influence in terms of components. Specifically, classical Granger causality only considers the contribution of the external driving force, i.e., the ratio between the variances of causal signal and noise, but ignores the structure of the system, i.e., the self causal effect and the interactions between different driving forces. Componential Granger causality is an improvement of the classical definition which assesses all causal contributions with a unitary measure, making the causal strengths in different systems comparable. In particular, componential Granger causality takes into account the previous history of $x$ on $x$, and the effect of noise on $x$, and the interactions between $y$ and $x$, in addition to the effect considered in classical Granger analysis of the past history of $y$ on $x$. The natural introduction of positive and negative interaction between $x$ and $y$ is also a particularly interesting feature of componential Granger causality. This can be regarded as an extension of the original idea of Wiener and Granger that causality can be detected by inspecting whether the past information of a signal (i.e. the effect of $y$ on $x$) can help reduce the prediction error. Here, with the componential definition, we can move one step forward to measure how the driving signal will affect the system, when causal influences are present. A positive causality implies that the input signal $y$ reduces the prediction error of the target signal $x$ and weakens the contribution of the noise term to the total causality. On the other hand, a negative causality value means the causal input (from $y$ to $x$) also reduces the prediction error but the negative effect of the interaction between $x$ and $y$ exceeds the contribution of the past information of $y$. More precisely, the cross term of componential Granger causality $F^c_{y\to x}$ defined in Eq. (6) can become negative if half of Component 3 when negative is larger than Component 1. This componential approach provides us with more information about the meaning of the causality value estimated from the data than is provided by classical Granger causality. Componential Granger causality also has the advantage that its terms sum to 1, so that different systems can be more directly compared. Another advantage of componential Granger causality is that it is less affected by differences in the variance of the noise. Classical Granger causality considers only the effect of $y$ on $x$, and expresses this relative to the noise. This makes the classical measure very sensitive to the variance of the noise.

The componential Granger causality analysis method described here can be applied to very many types of data, including neurophysiological data such as single or multi-neuronal recordings, optical imaging, fMRI, magnetoencephalography, local field potentials, and beyond neuroscience also to any possibly causal system where there are time series of data from two or many sources. We suggest that componential Granger causality analysis may therefore have applications in many fields, including economics, psychology, sociology, animal behavior, and climate change.

Componential Granger causality can also be linked to dynamic Bayesian network inference (Zou and Feng, 2009). To detect whether a process $X$ has a causal influence on $Y$, consider the simplest model where the network consists of only two nodes representing two random variables $X_{t-1}$ and $Y_t$. Suppose both $X_{t-1}$ and $Y_t$ are Gaussian distributed satisfying $X_{t-1}\sim\mathcal{N}(0,\sigma_1^2)$ and $Y_t\sim\mathcal{N}(0,\sigma_2^2)$ with correlation coefficient $\rho$ where $\mathcal{N}(\mu,\sigma^2)$ denotes the normal distribution with mean $\mu$ and variance $\sigma^2$. Under these assumptions and the dynamic Bayesian framework, the question of whether $X$ has a causal influence on $Y$ can be converted into whether $Y_t$ itself, or $Y_t\left|X_{t-1}\sim\mathcal{N}\left(\rho\frac{\sigma_2}{\sigma_1}X_{t-1},(1-\rho^2)\sigma_2^2\right)\right.$ can better explain the observed data. Suppose the process $X$ and $Y$ admit the linear structure $Y_t=aX_{t-1}+\varepsilon_t$ simultaneously where $\varepsilon_t\sim\mathcal{N}(0,\sigma^2)$; it follows that $Y_t|X_{t-1}\sim\mathcal{N}(aX_{t-1},\sigma^2)$. Combining the above conclusions, we have $\rho\frac{\sigma_2}{\sigma_1}X_{t-1}=aX_{t-1}$ and $(1-\rho^2)\sigma_2^2=\sigma^2$. According to the componential definition of the Granger causality, the causal strength from $X$ to $Y$ can be measured as $F^c_{X\to Y}=\frac{\mathrm{var}[aX_{t-1}]}{\mathrm{var}[aX_{t-1}]+\mathrm{var}[\varepsilon_t]}=\rho^2$. If $a=0$, $X_{t-1}$ and $Y_t$ are independent which means the inclusion of $X_{t-1}$ will not improve the estimation of $Y_t$. If $\mathrm{var}[aX_{t-1}]/\mathrm{var}[\varepsilon_t]\to\infty$, $X_{t-1}$ and $Y_t$ are linearly correlated. A strong causal influence exists in this case.

We further note that dynamic causal modeling (DCM) (Friston, 2008) is another widely used approach to detect causal influences and effective connectivity, and is closely related to classical Granger causality (Ge et al., 2009). Componential Granger causality outperforms DCM in several aspects. The most significant aspect is a detailed decomposition of the causal effect from different sources which DCM fails to discriminate, especially the driving force from the noise and the interaction between different sources. DCM is also not well tailored for data without experimentally designed inputs, and requires a strong knowledge of how the observations are caused. As a phenomenological model, componential Granger causality has an easy implementation, and can be applied to any temporal recordings. The natural inclusion of delays in the model is also an attractive feature of componential Granger causality over DCM since delay is an omnipresent phenomenon in the real world.

In conclusion, we have described a new approach to the estimation of causality, componential Granger causality. It has advantages over classical Granger causality, including measurement of interaction effects; the effects of the past history of $x$ on $x$ as well as the past history of $y$ on $x$; and providing for all the components to sum to 1 so that different systems can be compared. We have illustrated how it can be applied and interpreted, by applying it to functional neuroimaging data. Componential Granger causality analysis showed that there is a top–down effect from different parts of the lateral prefrontal cortex that controls activity in the orbitofrontal cortex vs anterior insular primary taste cortex when attention is to the pleasantness vs intensity of taste sensory inputs. Finally, we showed that componential Granger causality can reveal evidence on the mechanisms involved, with in this case the top–down biasing influence on attention operating especially when there is already activity in the system being biased.

## Acknowledgments

T.G. is supported by the China Scholarship Council (CSC). F.G. was supported by the Gottlieb-Daimler- and Karl Benz-Foundation, and by the Oxford Centre for Computational Neuroscience. The functional neuroimaging investigation was performed at the Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB) at Oxford University. We warmly thank Dr T.E. Nichols of The University of Warwick Department of Statistics for advice on the PPI analyses.

## Appendix A

We provide further information about the definition of componential Granger causality and its rationale in this Appendix.

First, Component 3 in Eq. (4) has a multiplier of 2 for the covariance. The reason for this is that the contribution of the total variance (with second order statistics) of the past information to the current state takes the form:

$$\mathrm{var}\left[a_{xx}(\mathcal{L})x_{t-1}+a_{yx}(\mathcal{L})y_{t-1}\right] = \mathrm{var}[a_{xx}(\mathcal{L})x_{t-1}] + \mathrm{var}\left[a_{yx}(\mathcal{L})y_{t-1}\right] + 2\,\mathrm{cov}\left[a_{xx}(\mathcal{L})x_{t-1}, a_{yx}(\mathcal{L})y_{t-1}\right].$$

This formulation ensures that the terms sum to 1. (This helps with the comparison between different systems, as described in the main text.)

Second, the reason why when we want to attribute the total contribution to $x$ and $y$, we group half of the interaction term with the contribution from $x$ (the self term) and another half with the contribution from $y$ (the cross term) is as follows. A time series will not only have a pure contribution from itself but will also interact with other time series which through covariance terms will enhance or counteract the contribution of that time series. Therefore, our definitions of componential Granger causality in Eqs. (6) and (7) can be interpreted as the overall effect of the previous history of $x$ on $x$, or the previous history of $y$ on $x$, including the cross term in Eq. (2), the self term in Eq. (3), and the interaction term in Eq. (4). Inspection of Eqs. (4) and (6) shows that the weighting of the covariance term in Eq. (6) is 1, i.e. the same as that of the cross term. Correspondingly, inspection of Eqs. (4) and (7) shows that the weighting of the covariance term in Eq. (7) is 1, i.e., the same as that of the self term. This equal weighting of the variance and covariance terms is reasonable in that variances are normally taken to sum together to contribute to the overall variance.

A more sophisticated justification for this definition comes from the link between this definition and some physical concepts, in particular that an equal split of the interaction term in Eq. (4) between the cross term in Eq. (6) and the self term in Eq. (7) has a plausible physical background. More specifically, consider the following simple two-dimensional autoregressive (AR) model:

$$\begin{cases} x_t = a_{xx}x_{t-1} + a_{yx}y_{t-1} + \varepsilon_{xt} \\ y_t = a_{xy}x_{t-1} + a_{yy}y_{t-1} + \varepsilon_{yt} \end{cases}$$

where $\varepsilon_{xt}$ and $\varepsilon_{yt}$ are independent white noise process with variances $\sigma_x$ and $\sigma_y$ respectively. This AR model can be regarded as a *Markov process* since the states of the system at time $t$ only depend on the states at time $t-1$ but not before. In mathematics and statistical physics, a key concept for Markov processes is *detailed balance* (Kampen, 1992). A Markov process is said to have detailed balance if it satisfies:

$$\pi(s)P(s,s') = \pi(s')P(s',s)$$

for any state $s$ and $s'$ where $\pi(s)$ is the stationary probability density, $P(s,s')$ is the transition probability density from state $s$ to state $s'$. To be simple, detailed balance implies that, around any closed cycle of states, there is no net flow of probability. We have shown, with some mathematics and intensive computation that under Gaussian assumptions, detailed balance leads to:

$$SNR_y F^c_{x\to y} = SNR_x F^c_{y\to x}$$

where $SNR_x = \mathrm{var}(x)/\sigma_x$, $SNR_y = \mathrm{var}(y)/\sigma_y$ are the *signal-to-noise ratio* of $x$ and $y$ respectively. $F^c_{x\to y}$ and $F^c_{y\to x}$ are the componential Granger causality we defined in the present paper, with the equal weighting of the covariance term and the cross or self terms. Therefore, with our definition, the net flow of probability can be linked with the flow of causality. This elucidates the background and rationale of our definition.

Third, we address some of the advantages of the new componential Granger causality analysis that we introduce. The main reason that we define the componential Granger causality from $y$ to $x$ as the contribution from $y$ to $x$ (the cross term, Eq. (2)) plus half of the interaction term (Eq. (4)) [and, correspondingly, the componential Granger causality from $x$ to $x$ as the pure contribution from the past information of $x$ to $x$ (the self term, Eq. (3)) plus half of the interaction term (Eq. (4))] is that this provides a natural graph representation of the causal influences detected. When there are multiple time series under analysis, a causal network is easily constructed in the usual way with this definition. However, if we leave out the interaction term (Eq. (4)) (which reflects 2 times the covariance term as shown in Eq. (4)) throughout, some additional nodes and many more links are needed in the network. For example, consider two time series $x$ and $y$ under analysis. The usual way is to estimate the causal influence from $x$ to $y$ as well as $y$ to $x$. However, if we have an additional interaction term here, one more node [the interaction between $x$ and $y$ denoted as $(x, y)$], and two more possible links [$(x, y)$ to $x$ and $(x, y)$ to $y$] are needed. This makes the interpretation much more difficult. Hence, we believe that our definition provides clearer and more compact results, with a representation familiar to those interested in causal analysis.

Of course, sometimes it will be helpful to investigate the self and cross terms (Components 1 and 2 in Eqs. (2) and (3)) and the interactions (Component 3 in Eq. (4)) separately and see how each behaves. Indeed, in our case, the contribution of each Component is shown in Table 3. Note that when attention is paid to pleasantness, the top–down effects of both Components 1 (the cross term) and Component 3 (the interaction term) are significantly larger than the bottom–up effects (corresponding to p values 0.0007 and 0.0227 respectively). (Top–down refers in our case to a direction of effect from the prefrontal cortex to the orbitofrontal cortex.) However, when attention is paid to intensity, if we only look at the pure contributions (the cross term, Component 1), the top–down effect is larger, but not significantly larger, than the bottom–up effect. Only when the overall effects captured by Eq. (6) (cross-componential Granger causality) are considered do significant top–down effects become apparent and significant. This is a typical example which shows that the effect of interaction will enhance the pure contributions. Therefore, we suggest use the measure of overall effects (our definition of componential Granger causality) when a graph representation and a summarized interpretation are needed while checking each component separately when more detailed underlying mechanisms are studied.

## References


Bar, M., 2007. The proactive brain: using analogies and associations to generate predictions. Trends Cogn. Sci. 11, 280–289.

Beck, D.M., Kastner, S., 2009. Top–down and bottom–up mechanisms in biasing competition in the human brain. Vision Res. 49, 1154–1165.

Bengtsson, S.L., Haynes, J.D., Sakai, K., Buckley, M.J., Passingham, R.E., 2009. The representation of abstract task rules in the human prefrontal cortex. Cereb. Cortex 19, 1929–1936.

Bressler, S.L., Menon, V., 2010. Large-scale brain networks in cognition: emerging methods and principles. Trends Cogn. Sci. 14, 277–290.

Bressler, S.L., Seth, A.K., 2011. Wiener-Granger Causality: A well established methodology. NeuroImage 58, 323–329.

Bressler, S.L., Tang, W., Sylvester, C.M., Shulman, G.L., Corbetta, M., 2008. Top–down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. J. Neurosci. 28, 10056–10061.

Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J. Comput. Assist. Tomogr. 18, 192–205.

Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. 3, 201–215.

de Araujo, I.E.T., Kringelbach, M.L., Rolls, E.T., Hobden, P., 2003a. The representation of umami taste in the human brain. J. Neurophysiol. 90, 313–319.

de Araujo, I.E.T., Kringelbach, M.L., Rolls, E.T., McGlone, F., 2003b. Human cortical responses to water in the mouth, and the effects of thirst. J. Neurophysiol. 90, 1865–1876.

de Araujo, I.E.T., Rolls, E.T., Kringelbach, M.L., McGlone, F., Phillips, N., 2003c. Taste-olfactory convergence, and the representation of the pleasantness of flavour, in the human brain. Eur. J. Neurosci. 18, 2059–2068.
Deco, G., Rolls, E.T., 2005a. Attention, short-term memory, and action selection: a unifying theory. Prog. Neurobiol. 76, 236–256.
Deco, G., Rolls, E.T., 2005b. Neurodynamics of biased competition and co-operation for attention: a model with spiking neurons. J. Neurophysiol. 94, 295–313.
Deshpande, G., Hu, X., Lacey, S., Stilla, R., Sathian, K., 2010a. Object familiarity modulates effective connectivity during haptic shape perception. Neuroimage 49, 1991–2000.
Deshpande, G., Sathian, K., Hu, X., 2010b. Effect of hemodynamic variability on Granger causality analysis of fMRI. Neuroimage 52, 884–896.
Desimone, R., Duncan, J., 1995. Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. 18, 193–222.
Ding, M., Chen, Y., Bressler, S.L., 2006. Granger causality: basic theory and application to neuroscience. In: Schelter, B., Winterhalder, M., Timmer, J. (Eds.), Handbook of Time Series Snalysis. Wiley, Weinheim, pp. 437–460.
Downar, J., Crawley, A.P., Mikulis, D.J., Davis, K.D., 2000. A multimodal cortical network for the detection of changes in the sensory environment. Nat. Neurosci. 3, 277–283.
Friston, K., 2008. Hierarchical models in the brain. PLoS Comput. Biol. 4, e1000211.
Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E.T., Dolan, R.J., 1997. Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6, 218–229.
Friston, K.J., Glaser, D.E., Henson, R.N., Kiebel, S., Phillips, C., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging: applications. Neuroimage 16, 484–512.
Ge, T., Kendrick, K.M., Feng, J., 2009. A novel extended Granger Causal Model approach demonstrates brain hemispheric differences during face recognition learning. PLoS Comput. Biol. 5, e1000570.
Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15, 870–878.
Geweke, J., 1982. Measurement of linear-dependence and feedback between multiple time-series. J. Am. Stat. Assoc. 77, 304–313.
Geweke, J.F., 1984. Measures of conditional linear-dependence and feedback between time-series. J. Am. Stat. Assoc. 79, 907–915.
Gitelman, D.R., Penny, W.D., Ashburner, J., Friston, K.J., 2003. Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. Neuroimage 19, 200–207.
Gottfried, J.A., O'Doherty, J., Dolan, R.J., 2002. Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. J. Neurosci. 22, 10829–10837.
Gottfried, J.A., Winston, J.S., Dolan, R.J., 2006. Dissociable codes of odor quality and odorant structure in human piriform cortex. Neuron 49, 467–479.
Grabenhorst, F., Rolls, E.T., 2008. Selective attention to affective value alters how the brain processes taste stimuli. Eur. J. Neurosci. 27, 723–729.
Grabenhorst, F., Rolls, E.T., 2010. Attentional modulation of affective vs sensory processing: functional connectivity and a top–down biased activation theory of selective attention. J. Neurophysiol. 104, 1649–1660.
Grabenhorst, F., Rolls, E.T., 2011. Value, pleasure, and choice in the ventral prefrontal cortex. Trends Cogn. Sci. 15, 56–67.
Grabenhorst, F., Rolls, E.T., Bilderbeck, A., 2008. How cognition modulates affective responses to taste and flavor: top down influences on the orbitofrontal and pregenual cingulate cortices. Cereb. Cortex 18, 1549–1559.
Grabenhorst, F., Rolls, E.T., Parris, B.A., D'Souza, A., 2010. How the brain represents the reward value of fat in the mouth. Cereb. Cortex 20, 1082–1091.
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.
Green, W.H., 2002. Econometric Analysis, 5 ed. Prentice Hall, Upper Saddle River, NJ.
Guo, S.X., Ladroue, C., Feng, J.F., 2010. Granger causality: theory and applications. In: Feng, J.F., Fu, W.J., Sun, F.Z. (Eds.), Frontiers in Computational and Systems Biology. Springer-Verlag, pp. 83–111.
Haase, L., Cerf-Ducastel, B., Murphy, C., 2009. Cortical activation in response to pure taste stimuli during the physiological states of hunger and satiety. Neuroimage 44, 1008–1021.
Henson, R., 2003. Analysis of fMRI time series. In: Ashburner, J., Friston, K., Penny, W. (Eds.), Human Brain Function. Academic Press, London.
Hwang, K., Velanova, K., Luna, B., 2010. Strengthening of top–down frontal cognitive control networks underlying the development of inhibitory control: a functional magnetic resonance imaging effective connectivity study. J. Neurosci. 30, 15535–15545.
Jiao, Q., Lu, G., Zhang, Z., Zhong, Y., Wang, Z., Guo, Y., Li, K., Ding, M., Liu, Y., 2011. Granger causal influence predicts BOLD activity levels in the default mode network. Hum. Brain Mapp. 32, 154–161.
Kampen, N.G.v., 1992. Stochastic Processes in Physics and Chemistry. Elsevier, Amsterdam.
Kanwisher, N., Wojciulik, E., 2000. Visual attention: insights from brain imaging. Nat. Rev. Neurosci. 1, 91–100.
Kiebel, S.J., Poline, J.B., Friston, K.J., Holmes, A.P., Worsley, K.J., 1999. Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. NeuroImage 10, 756–766.
Klein, J.C., Rushworth, M.F., Behrens, T.E., Mackay, C.E., de Crespigny, A.J., D'Arceuil, H., Johansen-Berg, H., 2010. Topography of connections between human prefrontal cortex and mediodorsal thalamus studied with diffusion tractography. Neuroimage 51, 555–564.
Kouneiher, F., Charron, S., Koechlin, E., 2009. Motivation and cognitive control in the human prefrontal cortex. Nat. Neurosci. 12, 939–945.
Kringelbach, M.L., O'Doherty, J., Rolls, E.T., Andrews, C., 2003. Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. Cereb. Cortex 13, 1064–1071.
Luo, Q., Ge, T., Feng, J., 2011. Granger causality with signal-dependent noise. Neuroimage 57, 1422–1429.
McCabe, C., Rolls, E.T., 2007. Umami: a delicious flavor formed by convergence of taste and olfactory pathways in the human brain. Eur. J. Neurosci. 25, 1855–1864.
Mechelli, A., Price, C.J., Friston, K.J., Ishai, A., 2004. Where bottom–up meets top–down: neuronal interactions during perception and imagery. Cereb. Cortex 14, 1256–1265.
Nitschke, J.B., Dixon, G.E., Sarinopoulos, I., Short, S.J., Cohen, J.D., Smith, E.E., Kosslyn, S.M., Rose, R.M., Davidson, R.J., 2006. Altering expectancy dampens neural response to aversive taste in primary taste cortex. Nat. Neurosci. 9, 435–442.
O'Doherty, J., Rolls, E.T., Francis, S., Bowtell, R., McGlone, F., 2001. The representation of pleasant and aversive taste in the human brain. J. Neurophysiol. 85, 1315–1321.
O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., Dolan, R.J., 2003. Temporal difference models and reward-related learning in the human brain. Neuron 38, 329–337.
O'Doherty, J.P., Buchanan, T.W., Seymour, B., Dolan, R.J., 2006. Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. Neuron 49, 157–166.
Renart, A., Parga, N., Rolls, E.T., 1999. Backprojections in the cerebral cortex: implications for memory storage. Neural Comput. 11, 1349–1388.
Rolls, E.T., 2008. Memory, Attention, and Decision-Making: A Unifying Computational Neuroscience Approach. Oxford University Press, Oxford.
Rolls, E.T., Deco, G., 2002. Computational Neuroscience of Vision. Oxford University Press, Oxford.
Rolls, E.T., Grabenhorst, F., 2008. The orbitofrontal cortex and beyond: from affect to decision-making. Prog. Neurobiol. 86, 216–244.
Rolls, E.T., Kringelbach, M.L., de Araujo, I.E.T., 2003. Different representations of pleasant and unpleasant odors in the human brain. Eur. J. Neurosci. 18, 695–703.
Rolls, E.T., Grabenhorst, F., Margot, C., da Silva, M.A.A.P., Velazco, M.I., 2008. Selective attention to affective value alters how the brain processes olfactory stimuli. J. Cogn. Neurosci. 20, 1815–1826.
Rossi, A.F., Pessoa, L., Desimone, R., Ungerleider, L.G., 2009. The prefrontal cortex and the executive control of attention. Exp. Brain Res. 192, 489–497.
Sakai, K., Passingham, R.E., 2003. Prefrontal interactions reflect future task operations. Nat. Neurosci. 6, 75–81.
Sakai, K., Passingham, R.E., 2006. Prefrontal set activity predicts rule-specific neural processing during subsequent cognitive performance. J. Neurosci. 26, 1211–1218.
Schippers, M.B., Keysers, C., 2011. Mapping the flow of information within the putative mirror neuron system during gesture observation. Neuroimage 57, 37–44.
Schippers, M.B., Roebroeck, A., Renken, R., Nanetti, L., Keysers, C., 2010. Mapping the information flow from one brain to another during gestural communication. Proc. Natl. Acad. Sci. U.S.A. 107, 9388–9393.
Schippers, M.B., Renken, R., Keysers, C., 2011. The effect of intra- and inter-subject variability of hemodynamic responses on group level Granger causality analyses. Neuroimage 57, 22–36.
Schoenfeld, M.A., Neuer, G., Tempelmann, C., Schussler, K., Noesselt, T., Hopf, J.M., Heinze, H.J., 2004. Functional magnetic resonance tomography correlates of taste perception in the human primary taste cortex. Neuroscience 127, 347–353.
Sims, C.A., 1972. Money, income, and causality. Am. Econ. Rev. 62, 540–552.
Small, D.M., Gregory, M.D., Mak, Y.E., Gitelman, D., Mesulam, M.M., Parrish, T., 2003. Dissociation of neural representation of intensity and affective valuation in human gustation. Neuron 39, 701–711.
Veldhuizen, M.G., Bender, G., Constable, R.T., Small, D.M., 2007. Trying to detect taste in a tasteless solution: modulation of early gustatory cortex by attention to taste. Chem. Senses 32, 569–581.
Wiener, N., 1956. The theory of prediction. In: Beckenbach, E. (Ed.), Modern Mathematics for Engineers. McGraw-Hill, New York.
Wilson, J.L., Jenkinson, M., Araujo, I.E.T., Kringelbach, M.L., Rolls, E.T., Jezzard, P., 2002. Fast, fully automated global and local magnetic field optimisation for fMRI of the human brain. Neuroimage 17, 967–976.
Yacubian, J., Glascher, J., Schroeder, K., Sommer, T., Braus, D.F., Buchel, C., 2006. Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. J. Neurosci. 26, 9530–9537.
Zou, C.L., Feng, J.F., 2009. Granger causality vs. dynamic Bayesian network inference: a comparative study. BMC Bioinformatics 10, 122.