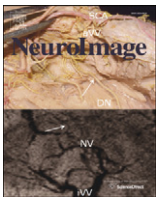
NeuroImage 72 (2013) 265–271

Contents lists available at SciVerse ScienceDirect

NeuroImage

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



Processing counterfactual and hypothetical conditionals: An fMRI investigation

Eugenia Kulakova [a,*], Markus Aichhorn [a], Matthias Schurz [a], Martin Kronbichler [a,b], Josef Perner [a]

[a] Department of Psychology and Centre for Neurocognitive Research, University of Salzburg, Salzburg, Austria
[b] Department of Neurology, Christian Doppler Clinic, Paracelsus Private Medical, University Salzburg, Austria

ARTICLE INFO

Article history:
Accepted 25 January 2013
Available online 4 February 2013

Keywords:
Counterfactual thinking
Conditionals
Subjunctive mood
Indicative mood
fMRI

ABSTRACT

Counterfactual thinking is ubiquitous in everyday life and an important aspect of cognition and emotion. Although counterfactual thought has been argued to differ from processing factual or hypothetical information, imaging data which elucidate these differences on a neural level are still scarce. We investigated the neural correlates of processing counterfactual sentences under visual and aural presentation. We compared conditionals in subjunctive mood which explicitly contradicted previously presented facts (i.e. counterfactuals) to conditionals framed in indicative mood which did not contradict factual world knowledge and thus conveyed a hypothetical supposition. Our results show activation in right occipital cortex (cuneus) and right basal ganglia (caudate nucleus) during counterfactual sentence processing. Importantly the occipital activation is not only present under visual presentation but also with purely auditory stimulus presentation, precluding a visual processing artifact. Thus our results can be interpreted as reflecting the fact that counterfactual conditionals pragmatically imply the relevance of keeping in mind both factual and supposed information whereas the hypothetical conditionals imply that real world information is irrelevant for processing the conditional and can be omitted. The need to sustain representations of factual and suppositional events during counterfactual sentence processing requires increased mental imagery and integration efforts. Our findings are compatible with predictions based on mental model theory.

Introduction

Counterfactuals describe events or states of the world that have not occurred and implicitly or explicitly contradict factual world knowledge. Counterfactual thinking is ubiquitous in everyday life and relevant for both cognition and emotion. People utter counterfactuals to indicate causal relations (Woodward, 2011) and to convey logical arguments (e.g. *reductio ad absurdum*). Counterfactual thinking is further necessary to express relief or regret (Kahneman and Miller, 1986), which is a motivational precondition for learning from mistakes and a form of emotional regulation (Markman et al., 1993). Some languages offer linguistic markers to clarify that an utterance is dealing with counterfactual suppositions: In English and German, subjunctive mood distinguishes counterfactual conditionals from those which do not imply real-world violations but rather have true (we might call them '*factuals*') or undetermined truth-values of their antecedents (further referred to as '*hypotheticals*') and which usually are stated in indicative mood.

Counterfactuals have been a topic in cognitive, social, and developmental psychology as well as linguistics for several decades and recently have received attention in psycho-linguistics (de Vega and Urrutia, 2012; de Vega et al., 2007; Ferguson, 2012; Ferguson and Sanford, 2008; Ferguson et al., 2008; Nieuwland, 2012, 2013; Nieuwland and Martin, 2012; Urrutia et al., 2012a, 2012b). In the cognitive domain of reasoning, mental model theory developed by Johnson-Laird and Byrne (1991) is probably the most prominent account that argues for a substantial difference between counterfactual and hypothetical conditional representation. Following Fillenbaum (1974), mental model theory proposes that subjects represent counterfactuals (e.g. *If it had rained then the street would be wet*) by constructing two distinct mental models: one for the literally uttered suppositional event (counterfactual: *rain* and *wet street*), and another one for the opposing factual event which is implicitly conveyed (factual: *no rain* and *dry street*) (Byrne, 2002). Hypotheticals, on the other hand (e.g. *If it rained then the street was wet*), only activate the suppositional model (*rain* and *wet street*), making no statement about factual events. Behavioral investigations support the notion of differential representations, showing that conditionals presented in past tense and subjunctive mood (i.e. counterfactuals) lead subjects to think of different implications and make different inferences than conditionals in past tense and indicative mood (i.e. hypotheticals). In particular, framing conditionals (If A then B) counterfactually increases the acceptance rate of negated suppositional events (¬A and ¬B) as well as the willingness to accept Modus Tollens (A → B, ¬B Therefore ¬A) but also the logically invalid deduction of Denying the Antecedent (A → B, ¬A Therefore ¬B). Both types of inferences are facilitated by the representation of factual events (¬A and ¬B), which is not triggered when

* Corresponding author at: Department of Psychology and Centre for Neurocognitive Research, University of Salzburg, Hellbrunner Strasse 34, 5020 Salzburg, Austria.
E-mail address: eugenia.kulakova@sbg.ac.at (E. Kulakova).

1053-8119
http://dx.doi.org/10.1016/j.neuroimage.2013.01.060

conditionals are presented in indicative mood (Byrne and Egan, 2004; Byrne and Tasso, 1999; Egan et al., 2009; Quelhas and Byrne, 2003; Thompson and Byrne, 2002).

However, results from paper and pencil tests do not clarify whether the construction of differential mental models occurs online during premise processing or with some delay in the course of implication and inference evaluation. Support for the view of an automatic and instantaneous activation of a factual model alongside the suppositional counterfactual model comes from reading-time investigations where subjects are not required to perform inferences but only read a text. Counterfactual but not hypothetical conditionals (If A then B) were shown to prime the opposing real-world model (¬A and ¬B), resulting in decreased reading times for subsequently stated events of ¬A and ¬B (Gomez-Veiga et al., 2010; Santamaria et al., 2005). Interestingly, this priming effect also worked the other way round: the antecedent of counterfactual conditionals (If A …) was read faster when subjects previously were introduced to the fact that ¬A (as opposed to the information that A). However, no difference in reading time occurred when the same information was conveyed in indicative mood, suggesting that factual world events were only activated by the counterfactual, but not by the hypothetical antecedent (Stewart et al., 2009).

Thus there is evidence indicating that the pragmatic implications of counterfactuals differ from those of hypotheticals in the sense that counterfactuals trigger an instantaneous co-activation of factual world knowledge alongside suppositional content. This knowledge affects online sentence processing, and in the long run results in different reasoning performance. In contrast, hypotheticals activate only one suppositional representation. On the basis of present evidence one would expect counterfactual and hypothetical conditionals to differ on a neural level during online sentence processing. In particular mental model theory, in which mental models are understood to be iconic representations of states or events, predicts differences in occipital brain regions where visual association processes and mental imagery take place (Knauff et al., 2003). Since counterfactuals are supposed to require more models than hypothetical conditionals these areas should be more strongly activated by counterfactual than by hypothetical conditionals.

However, no imaging study has targeted this prediction directly. Only one recent fMRI investigation offers a first glance into the neural basis of counterfactual sentence processing. Nieuwland (2012) used counterfactual conditionals to evaluate the effect of local context on truth value processing. He contrasted historical counterfactuals (*If N.A.S.A had not developed its Apollo Project, the first country to land on the moon would have been …*) to factual historical statements (*Because N.A.S.A has developed its Apollo Project, the first country to land on the moon was …*). The main effect of context (Counterfactual > Factual) showed increased bilateral activation in middle temporal gyri. Yet it is not clear from the reported stimulus material whether participants evaluated the presented historical counterfactual antecedents in respect to their factual world knowledge or only in a suppositional (hence hypothetical) way. Strong support for a hypothetical interpretation are the ERP results that Nieuwland and Martin (2012) obtained with the same stimulus material: there was no indication of a world-knowledge effect on the N400 during processing of contextually congruent (but factually false) target words embedded in counterfactual conditionals. This result is at odds with prior ERP investigations which reported influences of real-world knowledge on the N400 in counterfactual local contexts using stimulus material with strong real world salience (*If cats were vegetarian …*) (Ferguson et al., 2008; but see Nieuwland, 2013 for an alternative explanation of diverging results). Although truth value and context showed an interaction in neural activity during the presentation of the critical word at the conclusion in the fMRI-study, it is still possible that subjects took account of their real-world knowledge only when the conclusion was presented, triggered by the violation of a potentially strong association of context and critical continuation which is stored and automatically accessed from general world knowledge (e.g. *first country on the moon — America*). If subjects did not consider their factual knowledge regarding the presented counterfactual antecedents, the context effect observed by Nieuwland (2012) would reflect the neural difference between factual and hypothetical statements, rather than the contrast between counterfactual and factual statements.

In any case, to bring out the neural activation specific to counterfactual thinking it would be preferable to not only contrast counterfactuals with factuals but also with hypotheticals. Hypothetical, similar to counterfactual conditionals provide a supposition rather than a factual event and therefore have matching ontological status. Moreover, hypothetical and counterfactual conditionals both have the form of *if–then* statements instead of using temporal conjunctions like *since* or *because*. The only remaining difference is that hypothetical conditionals lack the counterfactuals' (explicit or implicit) antagonism to factual events, the hallmark of counterfactuality.

The present study was designed to implement this specific comparison. We presented counterfactual and hypothetical conditionals which were preceded by a sentence describing the factual state of events. This rendered our counterfactuals explicitly *counter-to-fact* which had an important benefit: It ensured that subjects took into account factual world knowledge and did not interpret the counterfactuals in a hypothetical way, as may have been the case with the stimulus material of Nieuwland (2012). To hold propositional content identical we presented the same fact-sentence in both counterfactual and hypothetical conditions (*The motor is switched off today.*). The suppositional content of the following counterfactual antecedent (*If the motor had been switched on today, …*) coincided temporally with opposing factual world events, resulting in an explicit antagonism. The hypothetical conditional, however, stated the same proposition but for a different time (*If the motor was switched on yesterday, …*) thus avoiding contradiction and leaving the truth value of the antecedent undetermined. To rule out primary perceptual confounds caused by inevitable sentence-length differences between counterfactual and hypothetical conditionals (due to the additional use of auxiliary verbs in subjunctive mood), stimulus sentences were presented in two different modalities, visually and aurally.

Material and methods

Participants

The sample included 21 healthy, right-handed volunteers (11 female) aged between 19 and 32 years ($M = 24.2$, $SD = 5.7$) with normal or corrected-to-normal vision. All were native German speakers with no history of neurological disorders. The subjects were recruited via online advertisement and gave informed consent before scanning. They were paid 10€ for participation.

Stimuli

The experimental material consisted of written and spoken German sentences. As can be seen in Table 1, in the counterfactual (CF) condition the first sentence described a factual physical event taking place *today* or *yesterday* (*The motor is switched off today.*). The second clause was formulated as a conditional interrogative sentence in subjunctive mood. Its antecedent described the opposite event taking place at the same time (*If the motor had been switched on today, …*), followed by a physical consequence that could (or could not) follow from the antecedent (*… would it have burned fuel?*). Negation was either implemented by direct 'not'-introduction (e.g. *burning* vs. *not burning*) or by using antagonistic verbs or adverbs (e.g. *off* vs. *on*).

The hypothetical (HYP) trials were composed of a fact-sentence identical to the counterfactual condition (*The motor is switched off today.*). The subsequent conditional, however, was phrased in indicative

Table 1
Examples of experimental trials.

	German original	English translation
CF	Der Motor ist heute **aus**.	The motor is switched **off** today.
	Wenn der Motor heute **an** wäre, …	If the motor had been switched **on** today, …
	… würde er dann Treibstoff verbrauchen?	… would it have burned fuel?
HYP	Der Motor ist heute **aus**.	The motor is switched **off** today.
	Wenn der Motor gestern **an** war, …	If the motor was switched **on** yesterday, …
	… hat er dann Treibstoff verbraucht?	… did it burn fuel?

Counterfactual (CF) and hypothetical (HYP) conditions. Bold and underlined font are used for clarity but were not used during stimulus presentation.

mood and described the opposite event taking place in a different time compared to the first sentence (either *today/yesterday* or *yesterday/today*). The hypothetically supposed antecedent (*If the motor was switched on yesterday, …*) was followed by a consequent indicating the same possible (or impossible) outcome as in the counterfactual condition (*… did it burn fuel?*). The temporal alternation allowed us to closely match propositional content of CF and HYP stimuli. However, it also introduced the confounding factor of temporal change in HYP. If present, we expected the potential process of adapting to temporal change to be reflected in the contrast HYP > CF only, thus not affecting our contrast of interest (CF > HYP).

In order to allow simple rating with no demand of context-specific knowledge, the conditionals described stable physical (*piano key pressed — tone generated*), technical (*cooker on — stove top hot*), and chemical (*noodles cooked — become soft*) relations which are available from general knowledge. Note that the acceptability of the relations was not affected by our manipulation. The only difference between conditions was the explicit contradiction between fact-sentence and antecedent in the CF condition which was missing in HYP conditions. This ensured that reasoning (i.e. conclusion verification) demands remained identical in both conditions. No personal or self-referring contents were presented and even animate agents were left out to keep the propositional content as simple as possible and exclude spontaneous social processes.

The pseudo-randomized experimental trials were interspersed with filler trials which consisted of the same first sentence as the counterfactual and hypothetical conditions (*The motor is switched off today.*), a conditionalized repetition of the introduced event (*If the motor is switched off today, …*), and the same question as both experimental conditions (*… does it burn fuel?*). This third condition was added to obscure the design from the subjects and prevent them from using simple implicit strategies to solve the task (e.g. using the temporal marker as a cue to condition), thus ensuring their attention during sentences processing.

In total 54 different themes were used to construct the counterfactual, hypothetical, and filler trials. Each theme was presented once per modality, therefore twice per subject, but always in different conditions within a subject.

As illustrated in Fig. 1, the durations of presentation were 3000 ms for the first sentence, 5000 ms for the antecedent, and 3500 ms for the consequent of the conditional interrogative sentence. In the visual condition the three parts were presented successively, preventing possible rereading of preceding information. Auditory presentation onsets coincided with those of the visual presentation, but the duration of spoken sentences showed a natural variation due to different lengths of the clauses. After the consequent which was framed as a question, two letters (*J* and *N* corresponding to *yes* and *no*) appeared for 1500 ms, prompting the subject to answer. Each trial was followed by a fixation-cross presented for 500 ms. Stimulus delivery and timing were controlled with Presentation (Neurobehavioral Systems, Albany, CA, USA). Three counterbalanced stimulus lists were employed in each of which trials of every condition were pseudo-randomly mixed

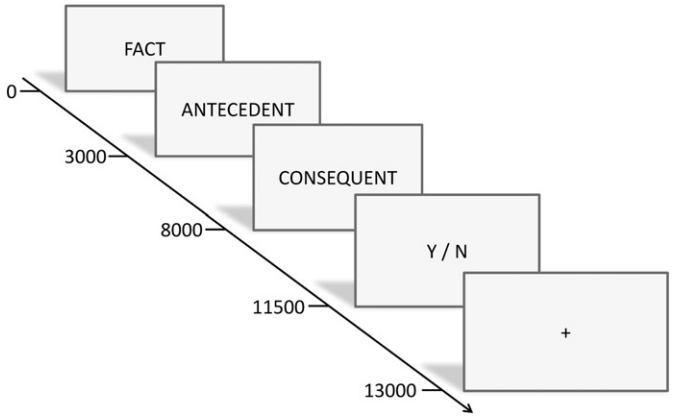


Fig. 1. Schematic illustration of stimulus onsets for both modalities. Axis indicates time in milliseconds.

with filler and baseline trials, thus balancing the succession order of identical trial types.

Procedure

The subjects were familiarized to the task prior to scanning. They were instructed to read the visually presented sentences carefully (or to listen to them in the auditory condition) and to answer the questions with *yes* or *no* after the letters *J* and *N* would have appeared on the screen. Subjects were encouraged to rely on their common sense and world knowledge instead of logical or probabilistic considerations. The visual stimuli were projected on a monitor located behind the head and could be seen through a mirror apparatus on top of the head coil. The auditory stimuli were presented via MRI-compatible pneumatic earphones. A fixation cross was presented on the screen during auditory stimuli delivery. Behavioral responses were collected with a two-button response box which the subjects were holding in their right hand and pressing with two different fingers (index finger for affirmation and middle finger for rejection). Positive and negative answers were balanced within and between conditions to prevent the use of low-level strategies and balance confounds of affirmation/negation and finger use.

Design

A categorical design with two main conditions (counterfactual, hypothetical) was employed for visual and auditory modality, resulting in a 2 × 2 within subject design. Functional imaging was subdivided into four sessions, two consecutive sessions per modality. Modality order was counterbalanced across subjects. Each session included 36 trials, 9 per condition (counterfactual, hypothetical), 9 filler, and 9 resting baseline trials where only a fixation cross was displayed for the length of a trial and which were used as an implicit baseline. Per subject, each of the 54 themes was presented once within a modality, but in different conditions (counterfactual, hypothetical, or filler). A single trial lasted 13 s, one session lasted 8.1 min and the whole functional experiment took 32.4 min to acquire.

fMRI data acquisition

Functional and structural imaging was acquired with a Siemens 3 Tesla Tim-Trio Scanner, located at Christian-Doppler-Clinic, Salzburg. Functional images sensitive to the BOLD contrast were obtained with a T2*-weighted gradient echo-planar imaging (EPI) sequence (TR = 2250 ms; TE = 30 ms; matrix size = 64 × 64; voxel size = 3 × 3 × 3 mm; slice gap 0.3 mm; FOV = 192 × 192 mm; flip angle = 70°). 36 axial slices were descendingly imaged parallel to the bicommissural

(co-planar with AC–PC) line, covering 118.8 mm of the z-axis. Per subject, four sessions, each one comprising 225 whole head images including 6 dummy scans at the beginning, were acquired. In addition to functional scanning, a high-resolution structural scan (T1-weighted MP-RAGE sequence; TR = 2300 ms; TE = 2.81 ms; voxel size 1.2 × 1 × 1 mm; slice-thickness = 1.2 mm; FOV = 240 × 256 mm; 160 slices; flip angle = 9°) was acquired sagitally to facilitate normalization and localization of functional data.

fMRI data processing

The fMRI data were processed using Statistical Parametrical Mapping (SPM 8, Wellcome Department of Imaging Neuroscience, London, UK) software, implemented in a MATLAB 7.6 (Mathworks, Sherborn, MA) runtime environment. Preprocessing was carried out for every subject individually. First, all images were pre-co-registered to an SPM8 EPI template. Functional data were then realigned in order to correct movement artifacts. Distortions in EPI caused by magnetic field inhomogeneity were corrected by using field maps for unwarping. In the next steps, structural data were segmented into white and gray matter and skull structures were separated from brain tissue. These steps were conducted in order to facilitate the normalization of the structural image to standard MNI space (Montreal Neurological Institute, McGill, Montreal, Canada). Last steps included co-registration of the mean functional scan to the structural data to normalize them both to standard MNI space and smoothing the functional data with an 8 mm full-width at half-maximum Gaussian kernel.

The preprocessed data were analyzed using a general linear model (GLM) approach. Per subject, modality, and session, each trial (CF, HYP, and filler) was modeled as a boxcar function with the duration of 11.5 s and convolved with a synthetic hemodynamic response function. The response frame was modeled as an event of no interest. Movement parameters were entered into the design matrix as additional regressors. Low-frequency noise was removed by a filter with a cutoff of 128 s and serial correlations were taken into account using an autocorrelation model AR(1). On the individual level, contrasts for CF and HYP relative to implicit fixation baseline were computed separately per modality. We did not differentiate between unanswered, correctly and incorrectly answered trials since errors and misses were very rare.

Data were taken on the second level and underwent a random effects analysis to allow for population inference. We computed a repeated-measures ANOVA by means of a full-factorial design following the guidelines by Glascher and Gitelman (2008). We modeled two factors (modality and condition) with two (visual, auditory) and two (CF, HYP) levels, respectively. Whole brain results are reported using an uncorrected threshold of $p<0.001$ with subsequent FWE cluster level correction with $p<0.05$. ROI analysis was conducted with the eigenvariate-tool of SPM8 using a non-repeated 2 × 2 design, averaging the mean brain activity estimates of the clusters from the supra-modal whole-level analysis. Since the only purpose of the ROI analysis is to show that the reported effects are not driven by one modality, we see no problematic circularity in this procedure.

Results

Behavioral results

Accuracy

The overall accuracy was high with around 95% (see Table 2). That is a good indicator that subjects were attentive and understood the task. We compared accuracy between modalities (auditory and visual) and conditions (CF and HYP) by comparing subjects' percentage of hit-rates with a 2 × 2 repeated measures ANOVA. There were no main effects or interaction (modality $F(1, 20) = 0.25$, $p = .62$; condition $F(1, 20) = 1.46$, $p = .24$; interaction $F(1, 20) = 1.38$, $p = .26$). This lack of significance indicates that the conditions were similar in difficulty.

Table 2
Behavioral results.

Modality	Condition			
	CF		HYP	
	% Hit (*SD*)	RT (*SD*)	% Hit (*SD*)	RT (*SD*)
Visual	96 (5)	640 (93)	96 (5)	635 (115)
Auditory	94 (8)	612 (109)	97 (5)	575 (141)
Both	95 (6)	626 (101)	97 (4)	605 (130)

Mean accuracy (percentage of hits) and reaction time (in milliseconds) over subjects for counterfactual (CF) and hypothetical (HYP) conditions. Standard deviations are indicated in brackets.

Reaction times

Reaction times for correct responses differed between conditions ($F(1, 20) = 5.47$, $p = .03$) but not modalities ($F(1, 20) = 3.21$, $p = .08$), with no significant interaction ($F(1, 20) = 0.91$, $p = .35$). Since the reaction times were collected in response to the Y/N-cue to elongate consequent evaluation, they may not reflect the actual time which was needed to judge the acceptability of the conclusions but rather measure the reaction time to the visually presented response cue. Nevertheless, counterfactual elaboration slowed down reaction time to the presented cue.

Imaging results

The contrast of interest was the comparison between counterfactual and hypothetical conditions across modalities. Here we found significantly stronger activity in right occipital cortex (cuneus) and marginally significant differences in right basal ganglia (caudate nucleus) on the whole-brain level as the main effect of Counterfactual > Hypothetical (see Table 3 for details). Post hoc *t*-tests showed that activation of both clusters was significantly stronger in the CF than in the HYP condition for each modality (cuneus visual $t(1, 20) = 5.68$, $p<.001$; cuneus auditory $t(1, 20) = 2.94$, $p<.01$; caudate visual $t(1, 20) = 2.53$, $p<.05$; caudate auditory $t(1, 20) = 3.56$, $p<.01$). It is especially noteworthy that the effect in the cuneus cluster is also present during auditory stimulus delivery (see Fig. 2). This rules out that the result is driven by differential processes related to visual perception. Comparisons in the opposite direction (HYP > CF) did not reveal any significant clusters. The main effect of modality (Visual > Auditory) was found in a predominantly left lateralized network which is typically found in visual word and sentence processing studies (Schurz et al., 2010). Stronger activation was found in bilateral occipital and occipito-temporal areas, as well as bilateral inferior and superior parietal areas. A large left frontal cluster was present, including inferior frontal gyrus and precentral gyrus. In the right frontal cortex, relatively small areas of stronger activation for visual compared to auditory sentence processing were found. The contrast in the other direction (Auditory > Visual) revealed activations in bilateral auditory cortices as well as large portions of bilateral temporal cortices, mainly including middle and superior temporal

Table 3
Supra-modal whole brain activations for main effect of condition.

	Region	H	MNI coordinates			*N* voxel	*T*(1,60)	p_{FWE}
			x	y	z			
CF > HYP	Cuneus	R	18	−73	10	82	4.49	0.01
			9	−94	13			
	Caudate	R	21	5	22	52	3.90	0.06
			15	20	16			
HYP > CF	*No activated clusters*							

Counterfactual (CF) and hypothetical (HYP). Significant clusters are reported at $p<.001$, the last column indicating significance for FWE cluster level correction. H = hemisphere, N = number, R = right.

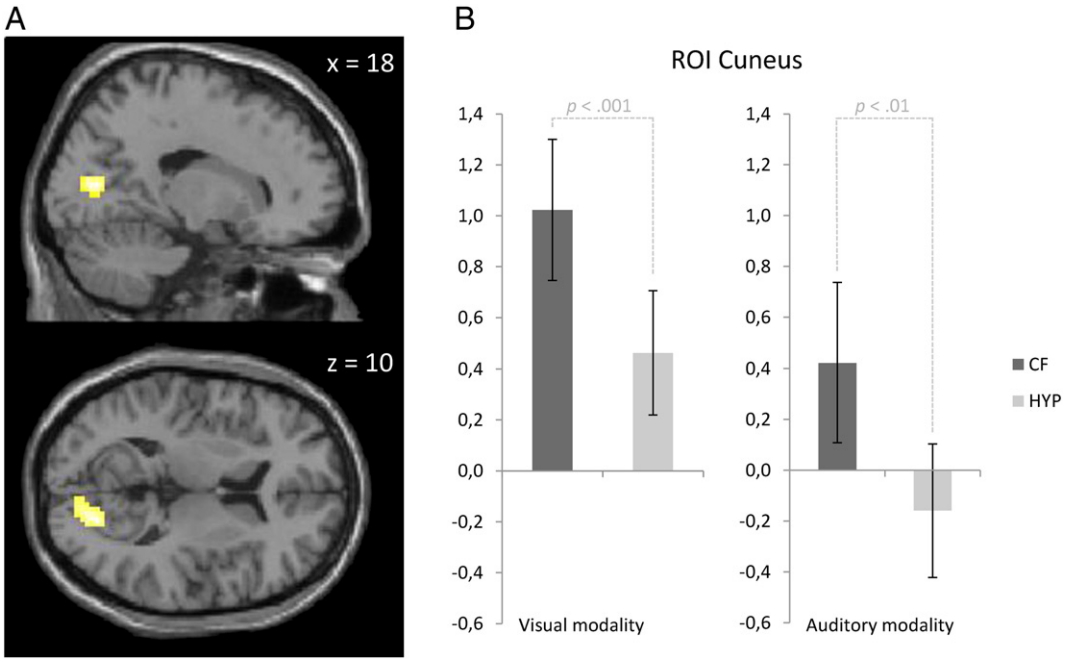


Fig. 2. Visualization of imaging results. A: Cuneus cluster from the supra-modal contrast CF > HYP projected on a single-subject structural image (Single T1 provided by the *standard SPM software package*); B: Mean brain activity estimates (given in arbitrary units) of the cuneus cluster, plotted individually per modality. Error bars indicate 2 SE which approximates a confidence interval of 95%.

gyri (see supplementary information for further details). No significant interaction effect was observed.

Discussion

In our task subjects were processing counterfactual conditionals in which they had to revise the consequence of assumed events. They were informed what was true in the real world, which made the antecedents of counterfactual conditionals explicitly contradicting stated facts. We contrasted this condition to hypothetical conditionals where subjects also knew the facts but were to think about a different hypothetical event which did not directly contradict the previously stated information. We argued that this condition provides the relevant contrast for identifying the neural correlates specifically involved in counterfactual thinking, since it controls for the *if–then* sentence structure and the ontological status of imagined events (both being suppositional, not factual), only differing from counterfactuals with respect to its antagonism to the factual event.

We found a neural difference between both conditions, namely right-lateralized activation in cuneus and (marginally) in the nucleus caudatus in the supra-modal contrast Counterfactual (CF) > Hypothetical (HYP). Stronger cuneus activation in CF was also present when stimuli were presented aurally, which is a strong indication that the observed difference builds on cognitive processes which differ between conditions but not modalities and therefore are not related to primary visual perception. Caudate activation was unexpected, but fits well with existing data on language comprehension and reasoning.

The role of occipital cortex

Occipital regions might be associated with the construction of a visual scene in absence of the appropriate external stimulus, a cognitive process which is called '*mental imagery*' (Thomas, 2010) and is a subcomponent of *scene construction*, the process of mentally generating and maintaining a complex and coherent scene (Hassabis et al., 2007). For our purposes we want to emphasize that these scenes may be intentionally created accompanied by a clear phenomenology of visualization (as the terminology suggests) but they could also be automatic processes triggered by counterfactual sentence comprehension lacking any conscious phenomenology of visualization. Summerfield et al. (2010) investigated scene construction by aurally presenting three-word descriptions of items which had to be incorporated into visually imagined scenes. Despite a non-visual presentation mode with closed eyes, occipital regions were activated bilaterally, more strongly during the imagination of the second scene element than during the first one and activation was even stronger when subjects were adding a third element to the imagined scene.

Two complementary explanations for higher imagery demands in the CF task are on hand. First, a visual load effect is possible. The broader and more complex the constructed scene becomes, the more occipital activation is required to capture it, akin to visual perception load. Alternatively, integration of a new visual element in the existing scene may be in some way more demanding and need additional resources. Indeed, both load and integration processes might be involved.

Occipital cortex and mental imagery load

In our stimulus material imagery load was matched by explicitly providing the same propositional information (*motor off, motor on, fuel burned*) in both experimental conditions. Therefore it seems at first sight implausible that CF needed more complex scenes. However, previously reported reasoning and text processing studies make it conceivable that the counterfactual antecedent reactivated the previously stated real world model and made it more salient to subjects whereas the hypothetical conditional signaled the move to a new and independent supposition and consequently did not cause such reactivation and amplification of factual events. In our contrast of interest the activations due to the representation of the suppositions canceled out, whereas the neural activity associated with the representation of the factual model did not. Our paradigm cannot show that the mere use of subjunctive mood triggers the representation of the actual state of the world since this information was explicitly provided in the stimuli. Nevertheless, the stronger cuneus activation indicates a more pronounced visual representation in the CF than

the HYP condition, which is well compatible with predictions based on mental model theory.

Occipital cortex and integration effect

The possibility of an effect of integration, however, remains possible and appears fairly plausible in the light of recent findings from the reasoning literature. Reasoning investigations use similar stimuli to our paradigm and occipital activations are frequently found but rarely discussed because frontal and parietal findings are more canonical. In his review Goel (2007) shows that more than half of reasoning studies yield occipital activation in differential contrasts. More evidence has accumulated since then and helps to pinpoint potential causes. Some occipital effects can be attributed to differences in reading behavior and sentence-length, especially when verbal reasoning content is presented visually. Other findings rule out primary perceptive differences through auditory presentation but are compatible with a mental imagery load interpretation (Just et al., 2004). This is especially likely when contrasts between concrete and abstract content (Knauff et al., 2003) or contrasts between spatial and conditional tasks (Knauff et al., 2002) are reported. However, there are some occipital findings from reasoning studies with verbal content which control for these variables and therefore need additional explanation. For instance, Prado et al. (2010) found stronger bilateral middle occipital activations during valid deduction from integrable arguments using Modus Tollens inference than during the processing of non-integrable control statements about shapes in a self-paced reasoning study with visual presentation mode. In integrable arguments both premises had one object in common (*There is not a circle. If there is a triangle then there is a circle.*) and therefore implied a conclusion (*There is no triangle.*). Non-integrable arguments had no common object in the second sentence (*There is not a circle. If there is a triangle then there is a diamond.*), which prevented any valid deduction. This integration effect cannot be reduced to perceptual differences between conditions, since sentences differed only in one word, which is also a reason why differential complexity of evoked images is unlikely to account for the observed differences. The integration effect is quite common in the reasoning literature. Goel and Dolan (2004) find bilateral occipital (and L putamen) contribution to inductive and deductive reasoning when compared to reading of three unrelated sentences in bilateral occipital cortex, and bilateral caudate for reasoning with concrete (*Karen is in front of Larry*) as well as abstract (*K is in front of L*) three-term relations (Goel and Dolan, 2001) compared to non-integrable statements.

One can conclude that integration of sentences in a single coherent and informative discourse model has often been associated with occipital activations. This supports the second interpretation of our results, namely effort due to successful integration. We also see that in most cases of integration, occipital activation occurs together with effects in the basal ganglia, and this is our only other main finding for the contrast CF > HYP.

Role of basal ganglia

Basal ganglia, especially in the left hemisphere, are an established part of the language-processing network as Friederici (2006) has emphasized. Basal ganglia seem to play a major role in controlled aspects of language processing when inhibition of preferred representations is required (Bornkessel-Schlesewsky and Schlesewsky, 2009). Longworth et al. (2005) put forward a non-language specific hypothesis of striatum contribution in language: suppression of competing alternatives during the late integrational processes in comprehension.

Basal ganglia and integration

It is intriguing that most reasoning investigations that contrasted integration with unrelated sentences show basal ganglia activations (Goel et al., 2000; Parsons and Osherson, 2001; Reverberi et al., 2010; in addition to the ones mentioned above). This is evidence that basal ganglia contribute to successful integration of linguistic information, be it within or between sentences. If such integration processes in sentence comprehension depend mostly on left basal ganglia, we hypothesize that right basal ganglia, as observed in our case, are contributing to the successful integration of two distinct and incompatible represented states of the world into one discourse model. This idea is inspired by the coarse coding-hypothesis (Beeman, 1998; Jung-Beeman, 2005) which assumes that the right hemisphere allows broader associations (for comprehending metaphors, deriving themes, understanding jokes), while the left hemisphere is more relevant for narrow, canonical meanings of words. Apparently, the integration of non-verbal, iconically presented spatial relations activates right rather than left basal ganglia, as found by Fangmeier et al. (2006) during premise integration when reasoning about the spatial relations of letters. Similar integration effort is needed in our task to keep in mind both mentally represented states (*motor on* and *motor off*).

Basal ganglia and truth evaluation

Another domain in which basal ganglia activation has been consistently reported is truth evaluation. Factually false sentences show increased right (Menenti et al., 2009) or bilateral caudate activation when compared to true statements (Nieuwland, 2012). It is interesting that the same effect is not present when counterfactual local context modifies truth values (Nieuwland, 2012), maybe because here both types of sentences (counterfactual true and counterfactual false) are false in respect to the real world. Since in our task counterfactual antecedents were false whereas hypothetical ones were not, the caudate finding is consistent with the literature. The truth-value contrast is furthermore comparable to the integration effect of incompatible models in the case of counterfactuals. In both cases the incoming propositional information cannot easily be incorporated in a single world model but needs alternative models to represent all information for a coherent discourse. So it is remarkable that representations of false statements and of counterfactual antecedents activate basal ganglia preferentially in the right hemisphere.

Basal ganglia and linguistic mood

On a lower level our results can be interpreted as effects of the differential use of subjunctive and indicative mood. As the manipulation of mood resulted in different structures of the verb-phrase, we might have found correlates of mood processing on a supra-modal neural level. This possibility would fit well the role of basal ganglia reported in language research, especially with syntactic complexity (Hagoort, 2003). However, its right lateralization is at odds with established linguistic theories and therefore fits better to our semantic integration interpretation. However, subjunctive mood and counterfactuality can only be dissociated with a paradigm that avoids the use of the subjunctive–indicative contrast for comparing CF and HYP conditionals, which is a difficult task to accomplish.

Conclusion

Our results show that counterfactual and hypothetical conditional processing differs in BOLD-responses in the right cuneus and marginally in the right caudate nucleus. These results cannot be trivially explained by basic perceptual processes or by differences in propositional content because they occur on a supra-modal level and between conditions closely matched in respect to conveyed content. Our favored explanation is that they reflect representational processes which differ between counterfactual and hypothetical conditionals. Counterfactual (subjunctive) conditionals imply the continued relevance of factual and suppositional information, suggesting that one should keep both in mind. In contrast, the hypothetical conditional suggests a switch from the foregoing real world information to a hypothetical scenario, suggesting that one can ignore the earlier information

for the new task at hand. The double representation of the counterfactual as proposed by mental model theory draws on increased implicit mental imagery resources as well as increased integration effort for creating a coherent discourse. Whether the increased unification effort is mainly driven by a propositional contradiction of factual and supposed events on a representational level (model interpretation) or reflects the use of subjunctive mood on a lower, syntactic level (language interpretation) remains to be clarified in further investigations.

Acknowledgments

The authors thank Eva Rafetseder for helpful comments on earlier versions of this manuscript and Steven Heywood for proof-reading. The first author of this article was financially supported by the Doctoral College "Imaging the Mind" of the Austrian Science Fund (FWF-W1233).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2013.01.060.

References

Beeman, M., 1998. Coarse semantic coding and discourse comprehension. In: Beeman, M., Chiarello, C. (Eds.), Right Hemisphere Language Comprehension: Perspectives from Cognitive Neuroscience. Erlbaum, Mahwah, New Jersey.
Bornkessel-Schlesewsky, I., Schlesewsky, M., 2009. Processing Syntax and Morphology: A Neurocognitive Perspective. Oxford University Press, USA.
Byrne, R.M., 2002. Mental models and counterfactual thoughts about what might have been. Trends Cogn. Sci. 6 (10), 426–431.
Byrne, R.M., Egan, S.M., 2004. Counterfactual and prefactual conditionals. Can. J. Exp. Psychol. 58 (2), 113–120.
Byrne, R.M., Tasso, A., 1999. Deductive reasoning with factual, possible, and counterfactual conditionals. Mem. Cognit. 27 (4), 726–740.
de Vega, M., Urrutia, M., 2012. Discourse updating after reading a counterfactual event. Psicológica 33, 157–173.
de Vega, M., Urrutia, M., Riffo, B., 2007. Canceling updating in the comprehension of counterfactuals embedded in narratives. Mem. Cognit. 35 (6), 1410–1421.
Egan, S.M., Garcia-Madruga, J.A., Byrne, R.M., 2009. Indicative and counterfactual 'only if' conditionals. Acta Psychol. 132 (3), 240–249.
Fangmeier, T., Knauff, M., Ruff, C.C., Sloutsky, V., 2006. FMRI evidence for a three-stage model of deductive reasoning. J. Cogn. Neurosci. 18 (3), 320–334.
Ferguson, H.J., 2012. Eye movements reveal rapid concurrent access to factual and counterfactual interpretations of the world. Q. J. Exp. Psychol. 65 (5), 939–961.
Ferguson, H.J., Sanford, A.J., 2008. Anomalies in real and counterfactual worlds: an eye-movement investigation. J. Mem. Lang. 58 (3), 609–626.
Ferguson, H.J., Sanford, A.J., Leuthold, H., 2008. Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. Brain Res. 1236, 113–125.
Fillenbaum, S., 1974. Information amplified: memory for counterfactual conditionals. J. Exp. Psychol. 102 (1), 44.
Friederici, A.D., 2006. What's in control of language? Nat. Neurosci. 9 (8), 991–992.
Glascher J., Gitelman D., 2008. Contrast weights in flexible factorial design with multiple groups of subjects. Unpublished tutorial. Retrieved from http://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind0803&L=SPM&P=R16629 (assessed 26.09.2012).
Goel, V., 2007. Anatomy of deductive reasoning. Trends Cogn. Sci. 11 (10), 435–441.
Goel, V., Dolan, R.J., 2001. Functional neuroanatomy of three-term relational reasoning. Neuropsychologia 39 (9), 901–909.
Goel, V., Dolan, R.J., 2004. Differential involvement of left prefrontal cortex in inductive and deductive reasoning. Cognition 93 (3), B109–B121.
Goel, V., Buchel, C., Frith, C., Dolan, R.J., 2000. Dissociation of mechanisms underlying syllogistic reasoning. NeuroImage 12 (5), 504–514.
Gomez-Veiga, I., Garcia-Madruga, J.A., Moreno-Rios, S., 2010. The interpretation of indicative and subjunctive concessives. Acta Psychol. 134 (2), 245–252.
Hagoort, P., 2003. How the brain solves the binding problem for language. NeuroImage 20 (1), S18–S29.
Hassabis, D., Kumaran, D., Maguire, E.A., 2007. Using imagination to understand the neural basis of episodic memory. J. Neurosci. 27 (52), 14365–14374.
Johnson-Laird, P.N., Byrne, R.M.J., 1991. Deduction. Lawrence Erlbaum Associates, Inc.
Jung-Beeman, M., 2005. Bilateral brain processes for comprehending natural language. Trends Cogn. Sci. 9 (11), 512–518.
Just, M.A., Newman, S.D., Keller, T.A., McEleney, A., Carpenter, P.A., 2004. Imagery in sentence comprehension: an fMRI study. NeuroImage 21 (1), 112–124.
Kahneman, D., Miller, D.T., 1986. Norm theory — comparing reality to its alternatives. Psychol. Rev. 93 (2), 136–153.
Knauff, M., Mulack, T., Kassubek, J., Salih, H.R., Greenlee, M.W., 2002. Spatial imagery in deductive reasoning: a functional MRI study. Cogn. Brain Res. 13 (2), 203–212.
Knauff, M., Fangmeier, T., Ruff, C.C., Johnson-Laird, P.N., 2003. Reasoning, models, and images: behavioral measures and cortical activity. J. Cogn. Neurosci. 15 (4), 559–573.
Longworth, C.E., Keenan, S.E., Barker, R.A., Marslen-Wilson, W.D., Tyler, L.K., 2005. The basal ganglia and rule-governed language use: evidence from vascular and degenerative conditions. Brain 128 (3), 584–596.
Markman, K.D., Gavanski, I., Sherman, S.J., McMullen, M.N., 1993. The mental simulation of better and worse possible worlds. J. Exp. Soc. Psychol. 29, 87–109.
Menenti, L., Petersson, K.M., Scheeringa, R., Hagoort, P., 2009. When elephants fly: differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. J. Cogn. Neurosci. 21 (12), 2358–2368.
Nieuwland, M.S., 2012. Establishing propositional truth-value in counterfactual and real-world contexts during sentence comprehension: differential sensitivity of the left and right inferior frontal gyri. NeuroImage 59 (4), 3433–3440.
Nieuwland, M.S., 2013. "If a lion could speak…": online sensitivity to propositional truth-value of unrealistic counterfactual sentences. J. Mem. Lang. 68 (1), 54–67.
Nieuwland, M.S., Martin, A.E., 2012. If the real world were irrelevant, so to speak: the role of propositional truth-value in counterfactual sentence comprehension. Cognition 122 (1), 102–109.
Parsons, L.M., Osherson, D., 2001. New evidence for distinct right and left brain systems for deductive versus probabilistic reasoning. Cereb. Cortex 11 (10), 954–965.
Prado, J., Van Der Henst, J.B., Noveck, I.A., 2010. Recomposing a fragmented literature: how conditional and relational arguments engage different neural systems for deductive reasoning. NeuroImage 51 (3), 1213–1221.
Quelhas, A.C., Byrne, R., 2003. Reasoning with deontic and counterfactual conditionals. Think. Reason. 9 (1), 43–65.
Reverberi, C., Cherubini, P., Frackowiak, R.S., Caltagirone, C., Paulesu, E., Macaluso, E., 2010. Conditional and syllogistic deductive tasks dissociate functionally during premise integration. Hum. Brain Mapp. 31 (9), 1430–1445.
Santamaria, C., Espino, O., Byrne, R.M., 2005. Counterfactual and semifactual conditionals prime alternative possibilities. J. Exp. Psychol. Learn. Mem. Cogn. 31 (5), 1149–1154.
Schurz, M., Sturm, D., Richlan, F., Kronbichler, M., Ladurner, G., Wimmer, H., 2010. A dual-route perspective on brain activation in response to visual words: evidence for a length by lexicality interaction in the visual word form area (VWFA). NeuroImage 49 (3), 2649–2661.
Stewart, A.J., Haigh, M., Kidd, E., 2009. An investigation into the online processing of counterfactual and indicative conditionals. Q. J. Exp. Psychol. 62 (11), 2113–2125.
Summerfield, J.J., Hassabis, D., Maguire, E.A., 2010. Differential engagement of brain regions within a 'core' network during scene construction. Neuropsychologia 48 (5), 1501–1509.
Thomas, N.J.T., 2010. Mental imagery. In: Zalta, E.N. (Ed.), Stanford Encyclopedia of Philosophy (Retrieved from http://plato.stanford.edu/entries/mental-imagery (assessed 26.09.2012)).
Thompson, V.A., Byrne, R.M., 2002. Reasoning counterfactually: making inferences about things that didn't happen. J. Exp. Psychol. Learn. Mem. Cogn. 28 (6), 1154–1170.
Urrutia, M., de Vega, M., Bastiaansen, M., 2012a. Understanding counterfactuals in discourse modulates ERP and oscillatory gamma rhythms in the EEG. Brain Res. 1455, 40–55.
Urrutia, M., Gennari, S.P., de Vega, M., 2012b. Counterfactuals in action: an fMRI study of counterfactual sentences describing physical effort. Neuropsychologia 50 (14), 3663–3672.
Woodward, J., 2011. Psychological studies of causal and counterfactual reasoning. In: Hoerl, C., McCormack, T., Beck, S. (Eds.), Understanding Counterfactuals, Understanding Causation: Issues in Philosophy and Psychology. Oxford University Press.