ISSN 1512-1801

MENTAL REPRESENTATION OF COUNTERFACTUAL THINKING: FROM ICONIC MINIMUM TO ABSTRACT MAXIMUM

Guillermo Macbeth[1,2] & Eugenia Razumiejczyk[1,2]

[1] National Scientific and Technical Research Council of Argentina (CONICET), Argentina
[2] Centre for Interdisciplinary Research in Values, Integration and Social Development at the Pontifical Catholic University of Argentina (UCA), Argentina

Corresponding Author: Guillermo E. Macbeth, Consejo Nacional de Investigaciones Científicas y Técnicas & Pontificia Universidad Católica Argentina, Centro de Investigación Interdisciplinaria en Valores, Integración y Desarrollo Social, Buenos Aires 249, E3100BQF, Paraná, Entre Ríos, Argentina. g.macbeth@conicet.gov.ar

Abstract

The aim of this contribution is to provide a state-of-the-art concerned with the mental representation of counterfactual thinking. Counterfactuals are defined from the perspective of cognitive psychology as reasoning processes that consider alternative situations to a factual situation. That is, counterfactual thoughts are conditional thoughts that negate a given fact. Therefore, the key problem of counterfactual mental representation is the representation of negation, its mental modeling and derived inferences. In this context, four prominent findings and three main open issues are revised. Our main conclusion states that counterfactual representation is a function of working memory load and probably operates on the basis of an abstraction gradient. That is, iconic representations might suffice for lower loads and abstract representations might be required for higher working memory loads. Suggestions for a research agenda on counterfactuals are presented. Such agenda is concerned with further theoretical developments and experimental adjustments.

Keywords

Reasoning-counterfactuals-representation-negation-iconic-abstraction-memory

Introduction

What would have happened if Germany and Japan won World War II? Philip K. Dick [1] wrote a classic novel based on this counterfactual thought. That is, a counterfactual is a conditional reasoning that assumes an alternative to a given fact [2], which can be actual or imaginary. The role of counterfactuals in human reasoning is concerned with explanation of the past, preparation for the future, imagination of causal relation, elaboration of emotion and intention, and decision making [3]. In the context of human thinking research, counterfactuals are important because they provide critical information about representation and inference. Particularly, counterfactuals are critical to understand negation [2,4].

Counterfactuals are conditional utterances like “if the flowers had been roses, the trees would have been orange trees” [2]. Imagine, for example, that this utterance appears in a conversation between a speaker and a hearer about a garden with flowers and trees. The mental representation of the hearer automatically includes the counterfactual and the presupposed facts:

Counterfactual:	roses	orange trees
Presupposed facts:	no roses	no orange tress

Such dual mental representation of counterfactuals has been found and extensively replicated in reasoning experiments [3]. As it can be seen in this example, the abstract representation of negation is required to produce counterfactuals. It is mandatory to generate the presupposed facts in mind. Even if the contextual information defines two possibilities like “lilies or roses”, and the presence of lilies is counterfactually derived from the denial of roses, the operation of negation requires abstraction. That is, negation implies abstraction. Hence, any psychological account of counterfactuals requires the description and explanation of abstract inferential processes. However, the most comprehensive and robust evidence-based current theory of human thinking makes

ISSN 1512-1801

emphasis on concrete components, iconicity, and the management of low loaded working memory processes [5]. That is, the Mental Models Theory or MMT [5,6,7] holds that mental representation proceeds by imaginary scenarios that are so iconic, simplified, and affirmative as possible. Mental models include only what is true and affirmative. Only a second stance process, the explicit mental modeling might flesh out what is false, negations and other complex components that remain implicit in the initial mental modeling [6].

These considerations suggest that the central problem of counterfactuals is the problem of negation, which is one possible variety of abstraction. Hence, a sound theory of counterfactuals requires an adequate account for negation as abstraction in mental representation and derived inference processes. Such advancement has been successfully started by the MMT. That is, the representation of abstract negation has been postulated for the mental modeling of counterfactuals as a tag to initial models. These annotated models are supported by coherent theory and robust experimental evidence. However, some issues can be detected in the current state of knowledge.

This contribution continues as follows. First, we review the four main discoveries concerned with the psychology of counterfactuals based on experimental evidence. Then, we outline three open research issues that will probably guide the near future of this research field. Finally, we discuss our findings concerned with the current state-of-the-art on counterfactual reasoning.

Main Discoveries Concerned With Counterfactuals: Dual Representation

The mental representation of counterfactual reasoning is a hard task because it requires the consideration of the presupposed facts and the counterfactual itself, that is, the alternative information explicitly included in the counterfactual utterance [2]. Its dual nature stems from negation, which is psychologically hard to account for [8]. Counterfactuals are also a variety of a conditional, which is associated to a complex psychological phenomenon that includes an explicit mental model and the mention of further implicit models in the context of the MMT [6]. That is, the indicative conditional of the form “if there are roses, then there are orange trees” can be represented by means of explicit and implicit mental models.

Explicit models	roses	orange trees
Implicit models	…	

That is, the three dots notation used in the context of the model theory [7] indicates that further models are detected, but not explicitly represented. Psychologically, indicative conditionals do not include all the possibilities stated by sentential logic, which includes:

Truth table	roses	orange trees
	no roses	orange trees
	no roses	no orange trees

The last two rows explicitly indicate that the negation of the antecedent generates true conditionals according to logic. However, such situation is only implicitly noted when building mental models up.

Hence, the complexity of counterfactuals for the MMT stems from its dual nature, which can be considered as the representation of two models. Furthermore, three additional research findings have been derived from such duality. These findings are concerned with facilitation effects, embodied negation contributions, and the inference-to-alternates effect.

Facilitation Effects

The evidence obtained from several experiments [4] suggests that counterfactuals are mentally represented by means of specific patterns. That is, counterfactual information and presupposed facts. Given such representation pattern, some facilitation effects have been observed. Counterfactuals such as “If there have been roses, the trees would have been orange trees” prime conjunctions like “no roses and no orange trees”. This priming can be understood as an acceleration effect in the processing of a conjunction after processing a counterfactual [3]. Several facilitation affects have been experimentally tested. Briefly, counterfactuals of the form “if there have been x, then there would have been y” facilitate conjunctions of the form “no x and no y” [4]. “x” and “y”

ISSN 1512-1801

refer to atomic propositions. Therefore, further atomic evidence is needed to argue in favor of such facilitation effects. Particularly, a direct evaluation of the priming effects of counterfactuals on the atomic components of the studied conjunctions is needed. That is, the acceleration in the processing of "no x and no y" is a compound of "no x" and "no y", which also have to be facilitated when compared to "x" and "y". This issue provides a specific item for the research agenda in counterfactual research.

Embodied Negation Contributions

Two fundamental contributions to the understanding of counterfactuals have been found in the context of embodied negation theory. One is concerned with the representation of presupposed facts in binary contexts and the other with a disembodiment effect of negation for action verbs. Embodied theories of cognition conjectures that mental representation is experiential, that is, necessary connected to sensorial modalities, which implies a strong link to lower abstraction level rather than a higher level of abstraction. Therefore, for binary contexts in which only two possibilities are available, the negation of one of them triggers the preference for the other possibility. The alternate is preferred instead of negation itself. That is, for contexts like "x or y", when further information brings "no x", then "y" is preferred instead of "no x". This might happen because "no x" is an abstraction and "y" is nearer to a concrete experience than any abstraction. However, such alternates preference can also be explained by the MMT as a particular case of an inference-to-alternates effect [2], that is, a consequence of representing annotated mental models.

A disembodiment effect has been observed when verbs that imply a body movement are included in negative utterances like the ones compatible with counterfactuals. That is, body action utterances that activate specific brain areas related to such processes are deactivated when such utterances are negated [9]. However, it is not clear how abstraction can be represented according to the embodied negation theories. Furthermore, such view of mental representation fails to predict the inference-to-alternates effect for multiple context counterfactuals, that is, situations that include alternatives like "x or y or z" instead of a dual context like "x or y" [2].

In sum, an embodied theoretical approach to counterfactuals needs further development and evidence concerned with the representation of abstract operations like negation and multiple context situations.

The Inference-To-Alternates-Effect

Contexts are critical to understand counterfactuals. Binary contexts like the ones that include two variables, "x or y", trigger different phenomena than multiple contexts like those including three or more variables like "x or y or z". A strong facilitation of the alternate has been found for binary contexts, while the preference for an abstract representation of negation has been observed for multiple contexts. That is, "no x" facilitates "y" in binary context, but the same negation promotes "no x" itself instead of "y" or "z" in multiple contexts [2]. An important conclusion can be derived from this finding. The abstract representation of negation implies a lower working memory load than the iconic representation of all the alternates, which are "y" and "z" in this example. Since the human mind prefers lower memory loads than higher ones, abstract representations are selected against multiple iconic representations.

Open Issues

The Threshold Issue

Human working memory is a very limited cognitive resource [6,7]. Since reasoning requires the intensive use of working memory, an important problem concerned with thinking is such limited condition [5]. Counterfactual thinking seems to operate according to a gradient of working memory load [2]. A lower use of such resources seems to trigger iconic representations, that is, sensorial or experiential information related to the topics studied by the embodied cognition theories [10]. A higher load of working memory seems to require the use of symbolic representations, e.g., a mental

ISSN 1512-1801

model with a tag of negation [11]. For example, instead of representing “no oranges” by means of representing “apples” or “grapes”, our mental modeling might require the representation of “no oranges”, that is “oranges” with a negation tag that operates symbolically [2]. In such case, a threshold issue becomes relevant [12]. How much information can hold the iconic processing when dealing with counterfactuals? Where is the limit in memory load that triggers the shift from iconic representation to symbolic representation? Is this all about individual differences or a psychophysics approach might be available [13] for specific tasks, materials, or domains? Is such threshold nomothetic or only idiographic? These questions require answers based on further experimental evidence.

The Linguistic Issue

It is well documented that reasoning with negation is a very difficult cognitive task [13,14,15,16]. However, such difficulty might stem from its representation, but also from prior linguistic issues [17]. That is, the experimental paradigm often applied to test hypotheses about counterfactuals typically uses compound sentences as primes and then requires inferences as targets related to such sentences. Many studies suggested that language processing difficulties might generate shallow responses [15,16]. That is, it has been observed perseverations, fast selections of response options that lack semantic processing, or simply a response that neglects the task instructions in reasoning experiments. Therefore, further evidence is required to discard in each experimental paradigm a linguistic explanation of counterfactual reasoning. That is, an adequate experimental paradigm is needed to reject objections concerned with the pure linguistic difficulty as responsible for the observed response patterns. One possible strategy to elaborate on this issue is to ovoid compound response options, which are typical in counterfactual experiments. That is, a prime constructed as a set of contextual sentences is followed by a prime constructed as a compound sentence connected by the disjunction operator “or”. Hence, the selection of a response option is mediated by the representation of such compound operator. For example, “oranges or apples” [4] has been often used as a response option. By the opposite, if the response is simplified by means of atomic sentences, for example “oranges”, then some linguistic issues can be solved. If “no oranges” as prime leads to a response pattern of “oranges” as target instead of “apples” when a reasoning task is involved, then the answer might be linguistically guided instead of being the result of a cognitive inference. Such shallow phenomenon can be understood as the counterfactual correlate to the matching bias response in the classic paradigm of the Wason Selection Task [7], which can be interpreted as a linguistic phenomenon rather than a reasoning phenomenon.

The Abstract Representation Issue

The Mental Models Theory of human thinking is based on imagery rather than logic [18]. That is, this modeling account assumes that our mental representation is iconic, or at least so iconic as possible rather than abstract. Therefore, abstraction remains somehow as a strange process. That is, we can imagine a world in which oranges are given, but we cannot clearly imagine a world where oranges are not given. We can instead imagine a world where apples are given, which is an implicit negation of oranges. The explicit representation of “no oranges” requires concrete replacements for oranges to avoid conflict with the iconicity of mental models. Hence, a tag for negation is required. That is, a lower abstraction representation must also include a higher abstraction annotation to include negation, which is an abstract function [8]. Moreover, such shift in the models’ theory that includes a tag for negation approaches to a rival account, the set of formal logic theories of human thinking that includes explicit mental representation of negation [19,20]. The later set of theories consider that the operation of negation is basic or fundamental. Negation is not a tag added to an iconic representation. It might be rather a primitive component of human thought. In sum, the inclusion of a negation tag in the mental models’ theory requires further theoretical elaboration based on recent experimental evidence [2,8,11,13].This issue needs to be elaborated on to preserve the internal consistency of the mental models’ theory, which is the more comprehensive and robust evidence-based account to describe and explain human thinking [2,6].

ISSN 1512-1801

Discussion

We argue that the state-of-the-art in the psychology of counterfactuals requires a research agenda based on four main discoveries and three open issues. It has been found that counterfactuals rely on: 1) a dual mental representation, 2) selective facilitation effects, 3) the iconicity of mental imagery studied by embodied theories of cognition, and 4) a cognitive inferential effect that prefers the alternate option in binary context, but not in multiple contexts. The main issues are concerned with: 1) a threshold between iconicity and abstraction, 2) linguistic restrictions that include syntactic, semantic and pragmatic components, and 3) the representation of abstraction in a model that relies heavily on iconicity.

We propose for such agenda both theoretical and experimental challenges [21]. Further theoretical elaboration is needed to clearly include the representation of abstraction in the Mental Models Theory of human thinking. The "tag" of negation added to an iconic representation seems to operate iconically, but also as an abstract operator. Therefore, further philosophical distinctions and precisions are required to define the epistemic status of such tag. Negation is central for counterfactuals. Hence, the conceptual elaboration of its representation is critical for the understanding of such duality. Similarly, further experimental design insights are needed to reject alternative linguistic explanations. In particular, a task design strategy that breaks the classic disjunction response option into its atomic components might solve several problems. Facilitation experiments based on disjunctions of the form "x or y" need to be replicated using "x" and "y" as separate, but not exclusive, response options.

The lack of the aforementioned theoretical elaboration might produce the atomization of the Mental Models Theory. That is, the problem of abstraction might generate separate theories to account for iconic processes by one side, and theories of abstraction by the other side. Similarly, the lack of the aforementioned experimental adjustments might produce confound effects in the interpretation of the evidence.

Conclusions

The inspection of the current state of knowledge concerning counterfactual research suggests that: 1) Counterfactual representation seems to be a function of working memory load. 2) It probably operates on the basis of an abstraction gradient. That is, iconic representations suffice for lower loads, but abstract representations might be required for higher working memory loads. 3) A specific research agenda can be proposed to promote the advancement of counterfactuals research. Both theoretical and experimental developments are required to understand representation and inference of counterfactual thinking.

ISSN 1512-1801

References

1. Dick, P. K. The Man in the High Caste, New York: Putnam, 1962
2. Espino, O. & Byrne, R. Thinking about the opposite of what is said: counterfactual conditionals and symbolic or alternate simulations of negation. Cognitive Science, 2018, 42(8), 2459-2501.
3. Byrne, R. M. J. Counterfactual thougth. Annual Review of Psychology, 2016, 67, 135-157.
4. Santamaría, C., Espino, O., & Byrne, R. Counterfactual and semifactual conditionals prime alternative possibilities. Journal of Experimental Psychology: Learning, Memory, and Cognition, 2005, 31(5), 1149-1154.
5. Johnson-Laird, P. N. Mental models and human reasoning. Proceedings of the National Academy of Sciences, 2010, 107(43), 18243-18250.
6. Johnson-Laird, P. N. How we reason, New York, NY: Oxford University Press, 2008.
7. Johnson-Laird, P. N. Mental models: Towards a cognitive science of language, inference, and consciousness, Cambridge, MA: Harvard University Press, 1983.
8. Khemlani, S., Orenes, I., & Johnson-Laird, P. N. Negation: A theory of its meaning, representation, and use. Journal of Cognitive Psychology, 2012, 24(5), 541-559.
9. Tettamanti, M., Manenti, R., Della Rosa P.A., Falini, A., Perani, D., et al. Negation in the brain: Modulating action representations. Neuroimage, 2008, 43, 358-367.
10. Mayo, R., Schul, Y., & Burnstein, E. "I am not guilty" vs "I am innocent": successful negation may depend on the schema used for its encoding. Journal of Experimental Social Psychology, 2004, 40(4), 433-449.
11. Orenes, I., Beltrán, D., & Santamaría, C. How negation is understood: Evidence from the visual world paradigm. Journal of Memory and Language, 2014, 74, 36–45.
12. Kahneman, D. Thinking fast and slow, New York, NY: Farrar, Strauss & Giroux, 2011.
13. Khemlani, S., Orenes, I., & Johnson-Laird, P. N. Negating compound sentences. In book: Building Bridges Across Cognitive Sciences Around the World. Miyake, N., Peebles, D., & Cooper, R. P (Eds.). Austin, Texas: Cognitive Science Society, 2012, 575-580.
14. Khemlani, S., Orenes, I., & Johnson-Laird, P. N. The negations of conjunctions, conditionals, and disjunctions. Acta Psychologica, 2014, 151, 1-7.
15. Macbeth, G., Crivello, M. C., Fioramonti, M. & Razumiejczyk, E. Chronometrical evidence supports the model theory of negation. Sage Open, 2017, 7(2), 1-8.
16. Macbeth, G., Razumiejczyk, E., Crivello, M. C., & Fernández, J. H. Gaze patterns of compound negation processing. Education Sciences and Psychology, 2017, 46(4), 3-10.
17. Horn, L. R. A natural history of negation (Rev. ed). Stanford, CA: CSLI Publications, 2001.
18. Johnson-Laird, P. N. Against logical form. Psychologia Belgica, 2010, 50(3&4), 193-221.
19. Rips, L. J. The psychology of proof. Cambridge, MA: The MIT Press, 1994.
20. Rips, L. J. Lines of thought. Central concepts in cognitive psychology. New York, NY: Oxford University Press, 2011.
21. Bonnefon, J. F. New ambitions for a new paradigm: Putting the psychology of reasoning at the service of humanity. Thinking & Reasoning, 2013, 19(3-4), 381-398.

Article received 2019-06-11