Interaction of polarity and truth value - A neural dynamic architecture of negation processing

Lea Kati (lea.kati@ini.rub.de)
Daniel Sabinasz (daniel.sabinasz@ini.rub.de)
Gregor Schöner (gregor.schoener@ini.rub.de)
Institute for Neural Computation, Ruhr-University Bochum, 44780 Bochum, Germany

Barbara Kaup (barbara.kaup@uni-tuebingen.de)
Psychological Institute, University of Tübingen, 72076 Tübingen, Germany

Abstract

We propose a neural dynamic architecture that models negation processing. The architecture receives a visual scene and a relational phrase like "The blue object is not to the right of the yellow object" or "The blue object is to the right of the green object" as input, and autonomously determines whether the phrase correctly describes the visual scene. The model is built out of empirically founded components for perceptually grounded cognition and constrained by neural principles. We demonstrate that the model can explain two commonly found reaction time effects: the negation effect in which reaction times are higher for negated than for affirmative phrases, and the polarity-by-truth-value interaction effect in which reaction times for false negated phrases are faster than those for true negated phrases whereas the opposite is true for affirmative phrases. The model is consistent with some aspects of the two-step simulation theory.

Keywords: negation effect; polarity-by-truth-value interaction; neural dynamics; dynamic field theory

Introduction

A key property of human language is, of course, that it can be used to describe the world around us. We perceptually ground a phrase by directing attention to objects or events in the world that the phrase refers to. Phrases may also be used to describe what is *not* the case in the world, however, expressed by negative rather than positive polarity. Moreover, as part of the perceptual grounding process, listeners may determine the truth value of phrases of either polarity.

Polarity entails non-trivial processing effects. Negated phrases are typically more difficult to process than affirmative phrases, as indicated for instance by longer reaction times and higher error rates when processing such phrases (Just & Carpenter, 1971; Clark & Chase, 1972; Wason, 1965). This finding is usually referred to as the negation effect. When phrases of varying polarity are presented at the same time as a visual scene they describe, a false negated phrase is often processed more rapidly than a true negated phrase whereas the opposite holds for affirmative phrases. This is referred to as the polarity-by-truth-value interaction (Kaup, Lüdtke, & Zwaan, 2005; Carpenter & Just, 1975; Chase & Clark, 1971)

The scientific stance of *grounded cognition*, as articulated, for instance, by Barsalou (2008), posits that language is intimately linked to perceptual representations, potentially leading up to perceptual symbols as units of representation. This supports attentional processes that select objects or events based on their properties and their relationships to other objects or events to achieve perceptual grounding.

Accounts for the actual neural networks that may represent elements of language in this way and for the actual neural processes of perceptual grounding are still under development (Holyoak & Hummel, 2000; Doumas & Hummel, 2012). We build on a dynamic field model of the perceptual grounding of relational phrases (Richter, Lins, & Schöner, 2021) that commits to a set of neural principles that constrain cognitive processes while being less concerned with the localization of cognitive functions in the networks of the brain.

Both the negation effect and the polarity-by-truth-value interaction provide an interesting challenge for a neural account of negation processing. Consequently, we build in this paper a neural process account for the role of polarity in the perceptual grounding of relational phrases and for how truth value of phrases of either polarity may be obtained.

Methods

The neural process account makes use of dynamic field theory (Schöner, Spencer, & the DFT Research Group, 2016), a mathematical framework to model cognitive processes consistent with neural principles. Psychological and behavioral effects are explained on the basis of the activation of neural populations, neural fields $u(x,t)$, which represent feature dimensions, x, such as visual space, or color by virtue of their forward connectivity from the sensory surfaces. Neural nodes represent categorical concepts. Neural activation evolves continuously in time (time scale, τ) not only driven by inputs, $s(x,t)$, from the sensory surfaces, but also shaped by strong current connectivity, $w(x-x')$, within the neural population:

$$\tau\dot{u}(x,t) = -u(x,t) + h + s(x,t) \, + \int g(u(x',t))w(x-x')dx'$$

where $g(u(x,t))$ is a sigmoid threshold function, and $h < 0$ the field's resting level. Sub-threshold activation becomes unstable for sufficiently strong localized input in the detection instability, leading to a switch to localized peaks of activation, the attractor states of the fields that serve as units of representation. When such peaks remain stable after inducing input is removed, they act as working memory. The decay of peaks in the reverse detection instability enables sequences of activation states that model cognitive processing. Neural nodes and fields can be combined into larger architectures, such as the one described in this paper.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Scenario

Both the negation effect and the polarity-by-truth-value interaction have been found in several studies (Kaup et al., 2005; Just & Carpenter, 1971; Clark & Chase, 1972). To examine the difference in processing phrases of different polarity and truth value, there are several paradigms, most of which involve the presentation of a phrase and an image. Then, participants are either asked to verify the sentence based on the image (as in Just & Carpenter, 1971; Clark & Chase, 1972) or to verify whether the objects mentioned in the phrase are present in the image (as in Kaup et al., 2005).

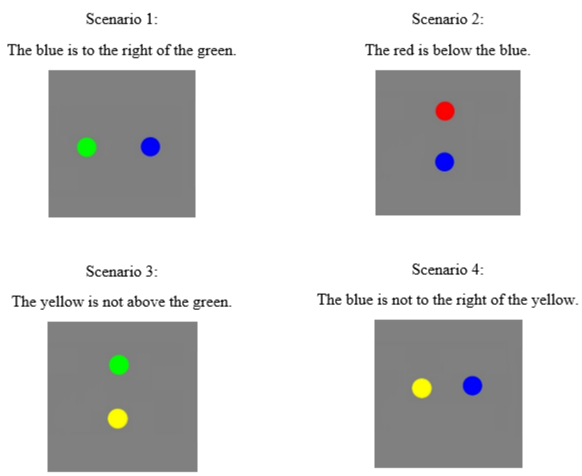

Figure 1: The model's task. The four scenarios represent the four possible polarity-truth-value combinations: true-affirmative, false-affirmative, true-negated, and false-negated.

For this model, a similar paradigm is used. It is presented with images that show two colored dots (red, blue, green, or yellow) in one of four relations to each other (above, below, to the right of, or to the left of), as well as with a phrase that describes the visual scene (see Figure 1). In relation to the image, the phrase is either true or false. The model's task is to first find the two objects in the image and to then identify the truth value of the phrase for four different conditions: a true affirmative phrase (Scenario 1), a false affirmative phrase (Scenario 2), a true negated phrase (Scenario 3), or a false negated phrase (Scenario 4).

Model

The goal is to provide a model within the DFT framework that may perceptually ground phrases with positive or negative polarity and to propose a neural dynamic mechanism for determining the truth value of such a phrase. We will then test if this model accounts for the negation effect and the polarity-by-truth-value interaction.

The neural architecture we propose is fundamentally one large dynamical system of coupled neural activation variables and fields. However, we can distinguish different sub-architectures in terms of their functional roles. The first and second functional sections of the model are simplifications of corresponding portions of the grounding model proposed by Richter et al. (2021).

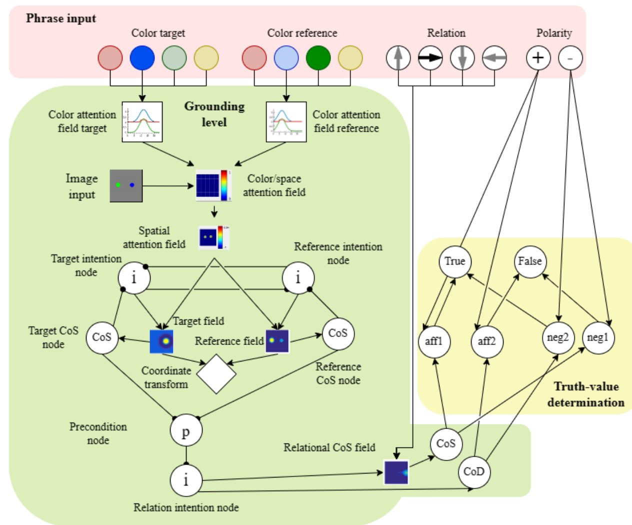

Figure 2: The model architecture divided into its three functional roles, phrase input (light-red), grounding level (green), and truth-value determination mechanism (yellow). The example scenario in this figure is Scenario 1 from Figure 1.

The phrase input (light-red section in Figure 2) consists of nodes, which represent discrete concepts. Color nodes for target and reference objects project onto the respective one-dimensional color attention fields, eliciting a peak in the corresponding area of the field. Both color attention fields, as well as the image, feed into a three-dimensional color/space attention field, with two space dimensions and one color dimension. The image input elicits sub-threshold bumps of activation in this field on colored objects; where depends on the saturation of the color. Peak formation then depends on the input from the color attention field, causing peaks only on objects of the given color. This field projects to the two-dimensional space field, which selectively allows for a peak on the object of the given color. The latter field projects to the target field and the reference field, creating a sub-threshold bump of activation there.

There are two processes: The target process and the reference process. They are organized by intention and condition-of-satisfaction (CoS) nodes (see Richter et al., 2021). The target and reference intention nodes suppress each other, meaning that only one of them can be active at a time. They provide input to the target or reference field, respectively, inducing peak formation. Once a target object or reference object has been found, meaning that a peak has been elicited in the corresponding field, the respective CoS node turns on. Via the inhibitory coupling, this deactivates the intention node, which in turn releases the intention node of the other process from inhibition, allowing that node to become active.

The target and reference fields feed into a neuronal transformation mechanism, which transforms the spatial location of the target into a coordinate system where the reference object is in the center. The result of this transformation feeds into a relational CoS field, which also receives input from the discrete relational nodes from the phrase input. It forms a peak if the relational concept given in the phrase matches the relative spatial relation of the objects.

This process of relational grounding is also organized. The intention node for the relational CoS field receives inhibition from a precondition node. This precondition node is suppressed once both the target and reference CoS nodes are activated. Thus, once both target and reference fields show an activation peak, the relational CoS field can receive input from the relation intention node. The corresponding relational CoS node gets activated once a peak has been formed in the relational CoS field. If no peak is formed after a certain amount of time, a condition-of-dissatisfaction (CoD) node is activated. This is enabled by virtue of excitatory input from the relational intention node and inhibitory input from the relational CoS node. The activation of either the CoS node or the CoD node reflects the decision the model makes whether the relation in the image matches the one in the phrase or does not.

The third functional section of the model is the truth-value determination mechanism (yellow section in Figure 2). The new concepts introduced here are the polarity of the phrase, positive polarity for an affirmative phrase and negative polarity for a negated phrase, and the two truth values, true and false. Two nodes represent the polarity of the phrase in the phrase input, and two nodes represent the truth values in the subarchitecture of the truth-value determination mechanism.

For this mechanism, the relational CoS node is projected onto the truth value *True* if the phrase has positive polarity and onto the truth value *False* if it has negative polarity. For the relational CoD node, these projections are crossed, meaning that if the phrase is affirmative, it projects onto the truth value *False* and if it is negated onto the truth value *True*. Neurally, this is possible via the four intermediate nodes *aff1, aff2, neg1, and neg2* (see Figure 2).

Results

For the experiment, we gave the model phrases and images, as shown in Figure 1. In the following, simulation demonstrations will be presented for the four possible scenarios (true-affirmative, false-affirmative, true-negated, and false-negated). These simulations are presented with respect to the time course of activation, which includes the time of grounding the objects and relations, as well as the time of truth-value determination.

Scenario 1: True affirmative phrase

Here, the model is given the phrase "The blue is to the right of the green," as well as the upper left image in Figure 1.

Figure 3 shows the time course of the grounding processes of the first task. The reference intention node (green bar) wins

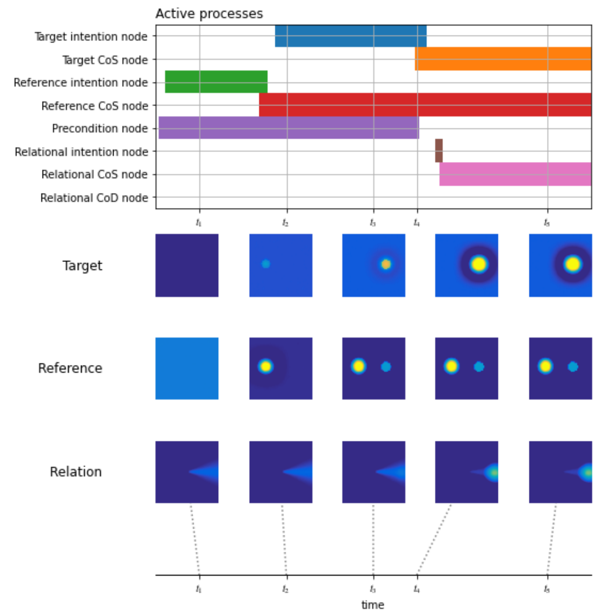

Figure 3: Processing of a true affirmative phrase. In the upper part of the figure, the nodes are represented for process organization. In the lower part of the figure, the fields are represented, where target object, reference object, and relation are grounded.

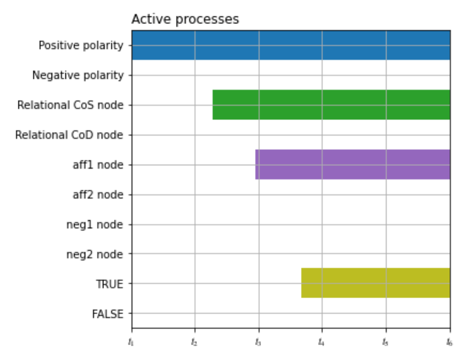

Figure 4: True affirmative phrase processing. Here, the process of truth value determination is represented.

the competition and initiates the grounding process of the reference object. By time point *t2*, there is a peak in the reference field, which causes the reference CoS node to turn on (red bar), inhibiting the reference intention node. Since the inhibition of the target intention node now stops, it turns on (blue bar), and by time point *t3*, a peak forms in the target field as well. This causes the target CoS node to be activated (orange bar). The precondition node (purple bar), which has been activated since the beginning now turns off as a result of inhibition from both the target and the reference CoS nodes. Therefore, the relational intention node (brown bar) turns on, and just after time point *t4*, a peak is elicited in the relational CoS field, activating the relational CoS node (pink bar).

In Figure 4, the time course of the truth value determination mechanism is shown. This time course overlaps with that of Figure 3 for the relational CoS and CoD node activations.

The node for positive polarity (blue bar) has been activated since the phrase was given. As mentioned in the previous section, the peak in the relational CoS field caused activation of the relational CoS node (green bar). Both activated nodes turn on the intermediate *aff1* node (purple bar), which then elicits activation of the node *True* (yellow bar).

Thus, the model identifies this affirmative phrase as being true with regard to the given image.

Scenario 2: False affirmative phrase

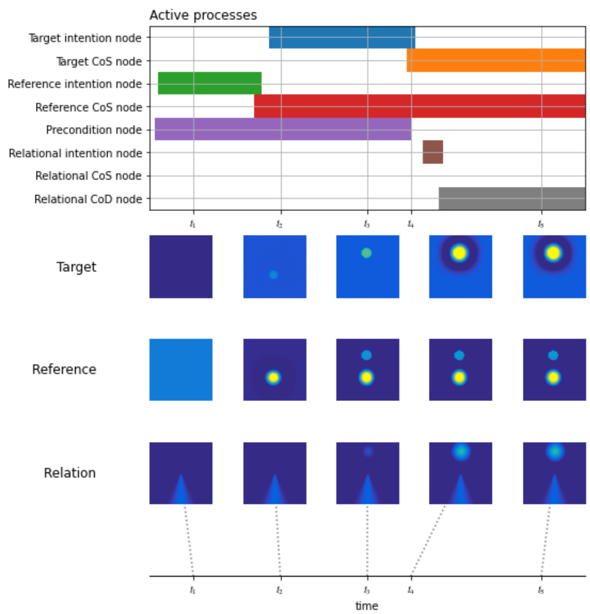

Figure 5: Processing of a false affirmative phrase. In the upper part of the figure, the nodes are represented for process organization. The lower part of the figure shows the fields, where target object, reference object, and relation are grounded.

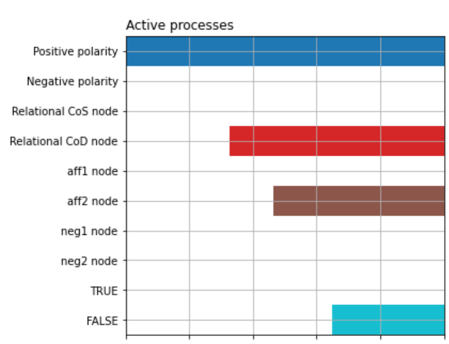

Figure 6: False affirmative phrase processing with respect to truth value determination.

For this scenario, the given phrase is “The red is below the blue” presented in the context of the upper right image from Figure 1.

Figure 5 shows the time course of the grounding processes of this second scenario. The processes are similar to those in the first scenario with a similar time course. First, the reference field shows a peak, then the target field. However, now there is no peak in the relational CoS field. Therefore, the relational CoD node is activated (grey bar).

Figure 6 shows the time course for the truth value determination mechanism for this scenario. The positive polarity node has been activated (blue bar) and together with the relational CoD node (red bar) provides input to the intermediate node *aff2* (brown bar). The *aff2* node activates the truth value node *False* (light-blue bar).

Therefore, the model identifies this affirmative phrase as false with regard to the given image.

Scenario 3: True negated phrase

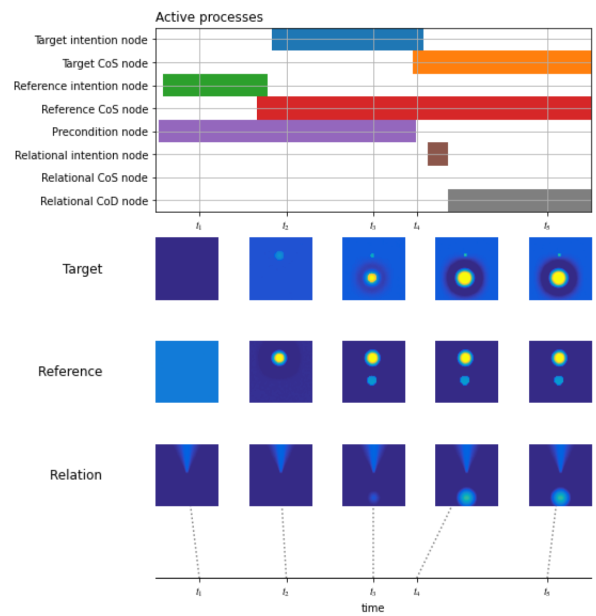

Figure 7: Processing of a true negated phrase. The upper part of the figure shows the nodes for process organization. In the lower part of the figure, the fields are represented, where target and reference objects, as well as relation are grounded.

For this scenario, the model is presented with the phrase “The yellow is not above the green,” as well as the lower left image in Figure 1.

Figure 7 shows the time course of the grounding processes of the third scenario. The processes are similar to those of the first two scenarios, also with regard to their time course. First, the reference object is found, and then the target object. As in the second scenario, there is no peak in the relational CoS field. Therefore, the relational CoD node is activated (grey bar).

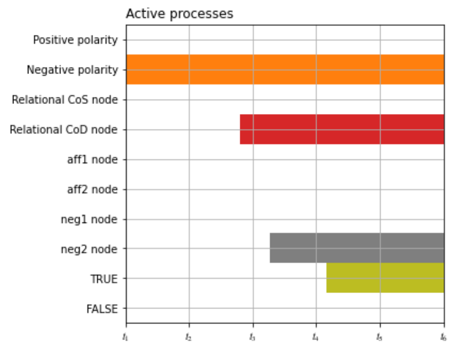

Figure 8: Node activation for truth value determination of the true negated phrase.

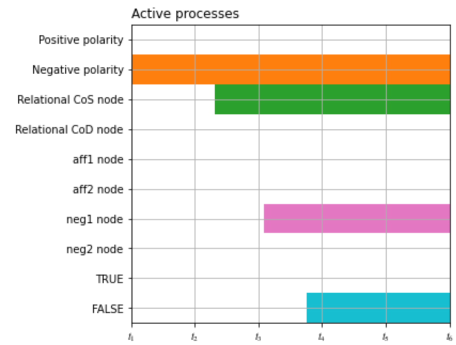

Figure 10: Node activation for truth value determination of the false negated phrase.

In Figure 8, the time course for the truth value determination mechanism for the third scenario is shown. In contrast to the first two scenarios, the negative polarity node has been activated from the phrase input (orange bar) and together with the relational CoD node (red bar) provides input to the intermediate node *neg2* (grey bar). The *neg2* node activates the truth value node *True* (yellow bar).

Therefore, the model identifies this negated phrase as true with regard to the given image.

Scenario 4: False negated phrase

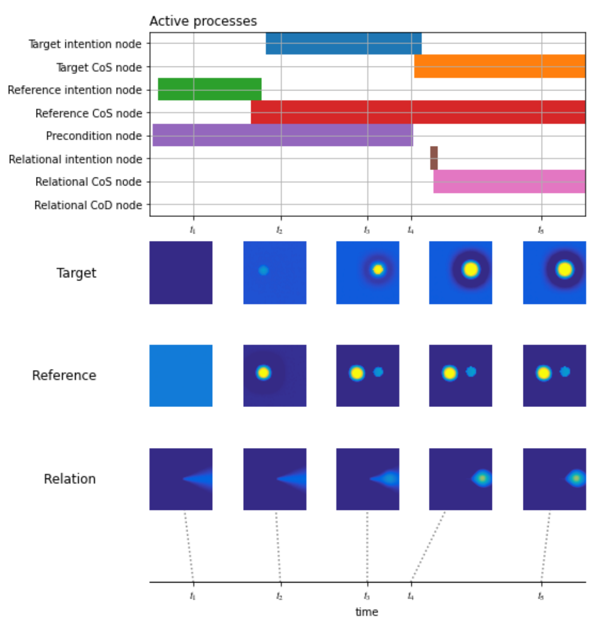

Figure 9: Processing of a false negated phrase. The upper part of the figure shows the nodes for process organization. In the lower part of the figure, the fields are represented, where target object, reference object, and relation are grounded.

For this scenario, the model is presented with the phrase "The blue is not to the right of the yellow," as well as the lower right image in Figure 1. In Figure 9, the time course of the grounding processes of the fourth scenario is shown. As in the first three scenarios, first, the reference object is found, and then the target object. As in the first scenario, there is a peak in the relational CoS field. Therefore, the relational CoS node is activated (pink bar).

Figure 8 shows the time course for the truth value determination mechanism for Scenario 4. As in the third scenario, the negative polarity node has been activated from the phrase input (orange bar) and together with the relational CoS node (green bar) provides input to the intermediate node *neg1* (pink bar). The *neg1* node activates the truth value node *False* (light-blue bar).

Thus, the model identifies this negated phrase as false with regard to the given image.

Reaction time effects

Additionally, this model shows the reaction time effects associated with negation comprehension, both the negation effect and the polarity-by-truth-value interaction, as can be seen in Table 1. Firstly, it can be seen that processing an affirmative true phrase takes 110 milliseconds less than a negated true phrase, showing the common negation effect. Secondly, processing a true negated phrase takes about 80 milliseconds more than a false negated phrase, while a true affirmative phrase takes about 110 milliseconds less than a false affirmative phrase. This illustrates the polarity-by-truth-value interaction. Thirdly, it can be noted that the negation effect disappears when the phrase is false, as a false negated phrase takes 80 milliseconds less to verify than a false affirmative phrase.

Table 1: The model's reaction times for all four scenarios.

	True	False
Affirmative	3.53 ms	3.64 ms
Negated	3.64 ms	3.56 ms

Discussion

We have thus presented, to the best of our knowledge, the first neural dynamic process model, a DFT architecture, that is capable of correctly identifying affirmative and negated phrases as true or false in relation to a given image. Additionally, the time course of activation in all four given scenarios accounts for the commonly found reaction time effects. Firstly, the model shows the negation effect, namely increased reaction times for processing true affirmative phrases versus true negated phrases. Secondly, the increased time to identify a negated phrase as true versus as false while an affirmative phrase takes longer to identify as false than as true is consistent with the polarity-by-truth-value interaction.

The two-step simulation theory, as suggested by Kaup et al. (2005), is a theoretical account of why the polarity-by-truth-value interaction occurs. It states that when confronted with negated phrases, humans first simulate the non-factual state, the affirmative version of the phrase, and only after some time adapt their mental simulation to the factual state, its negated version.

Our model is consistent with this theory in the sense that it first grounds the affirmative version of the phrase (the non-factual state). The truth value is determined only on the basis of the relational CoS or CoD node of the affirmative version of the phrase and its polarity. The polarity-by-truth-value interaction results from this account. Namely, if the phrase is negated, its affirmative version (or the non-factual state) is processed first, in this case perceptually grounded. If the negated phrase is false, its affirmative version will be true, meaning that the relational CoS node will turn on. Only then does the polarity play a role, activating the truth value *False* in the described crossing mechanism. If the negated phrase is true, however, its affirmative version will be false. The relational CoD node takes longer to turn on than the CoS node, while the crossing mechanism to the truth value *True* takes an equal amount of time. This way, the true negated phrase takes longer to verify than the false negated phrase. If the phrase is affirmative, however, the version of it to be grounded is already the factual state. Therefore, the activation of the relational CoS or CoD node directly projects onto the truth values *True* and *False*, respectively. Since the CoD node takes longer to be activated, the false affirmative phrase takes longer to process than the true affirmative phrase.

It is important to note that the model presented here differs in critical aspects from the processing model proposed by Clark and Chase (1972), which assumes that sentences and the outside world are both represented in a propositional representation. According to the model, these two representations are then compared constituent by constituent, with the response node turning from true to false and vice versa each time an inconsistency is being noticed. The negation-by-truth value interaction is explained by the number of times the response parameter has to be changed which is a very different explanation than the one presented here where the response time pattern is attributed to the fact that the model first grounds the affirmative version of the phrase.

A further question that poses itself is how the scope of the negation operator is determined. Here, we have assumed that the negation refers to the relation in the sentence, meaning that the relation is the part of the phrase that needs to be probed in order to verify the phrase or identify it as false. This is closely related to the concept of the "question under discussion" (QUD), as proposed by Roberts (2012). A phrase could have prosodic focus on another part of the phrase to indicate that this part should be checked. An example would be "The red is to the left of *the green*", in which the QUD would now be "Is it the green which the red is to the left of?". Here, instead of the relation, the reference object is in question. The model proposed in this paper may be extended to this wider set of tasks. In its current version, the relation, as the part under discussion, is grounded last. However, the order of grounding the three parts could be flexibly altered and adapted to the specific task.

Such an extension of this model seems particularly interesting as studies have shown that there are instances when the negation effect and the negation-by-truth value interaction effect decrease or disappear, such as when the context (1) invites the expectation of the non-factual state and is thus a felicitous context for a negation to appear, (2) is highly constraining, leading to high close probabilities, or (3) makes available specific alternatives to the non-factual state, such as when the context is felicitous (see Nieuwland & Kuperberg, 2008; Nieuwland, 2016; Orenes, Beltran, & Santamaria, 2016; Spychalska, Haase, Kontinen, & Werning, 2019; Kaup & Dudschig, 2020). For instance, the grounding of the affirmative version of the phrase does possibly not take place in situations with a negative QUD, such as when the negative sentence would read as "It is not the green that is to the left of the red" (Tian, Ferguson, & Breheny, 2016). This current model only accounts for cases where there are no contextual cues given and thus only represents the neural mechanism of negation processing without context. However, it does not exclude context. An extension of this model could, additionally to accounting for a QUD, also incorporate memory (Schöner et al., 2016) of the previous phrases. Such an extension could possibly account for the observed context effects in negation processing by providing the model with a basis for expecting the non-factual state and/or with plausible alternatives to the non-factual state.

Overall, we have shown a possible neural dynamic account of negation processing which not only shows the repeatedly observed negation effect but also produces the well-known negation-by-truth value interaction by employing one core mechanism suggested by the literature, namely the view that comprehenders ground the affirmative version of the phrase when processing a negative sentence.

References

Barsalou, L. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, *82*(1), 45–73.

Chase, W. G., & Clark, H. H. (1971). Semantics in the perception of verticality. *British Journal of Psychology*, *62*(3), 311–326.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*(3), 472–517.

Doumas, L. A. A., & Hummel, J. E. (2012). Computational Models of Higher Cognition. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (pp. 52–66). Oxford University Press.

Holyoak, K. J., & Hummel, J. E. (2000). The proper treatment of symbols in a connectionist architecture. In Deitrich, E & Markman, A (Eds.), *Cognitive dynamics: Conceptual change in humans and machines.* Cambridge, MA, USA: MIT Press.

Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 244–253.

Kaup, B., & Dudschig, C. (2020). Understanding negation: Issues in the processing of negation. In *The oxford handbook of negation* (pp. 789–812). Oxford, UK: Oxford University Press.

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2005). Effects of negation, truth value, and delay on picture recognition after reading affirmative and negative sentences. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 27).

Nieuwland, M. S. (2016). Quantification, prediction, and the online impact of sentence truth-value: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 316.

Nieuwland, M. S., & Kuperberg, G. R. (2008, December). When the Truth Is Not Too Hard to Handle: An Event-Related Potential Study on the Pragmatics of Negation. *Psychological Science*, *19*(12), 1213–1218. doi: 10.1111/j.1467-9280.2008.02226.x

Orenes, I., Beltran, D., & Santamaria, C. (2016). How negation is understood: Evidence from the visual world paradigm. *Journal of memory and language*, *74*, 36–45.

Richter, M., Lins, J., & Schöner, G. (2021). A neural dynamic model of the perceptual grounding of spatial and movement relations. *Cognitive Science*, *45*(10).

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*, 6–1.

Schöner, G., Spencer, J., & the DFT Research Group. (2016). *Dynamic thinking: A primer on dynamic field theory*. New York, NY: Oxford University Press.

Spychalska, M., Haase, V., Kontinen, J., & Werning, M. (2019). Processing of affirmation and negation in contexts with unique and multiple alternatives: Evidence from event-related potentials. *CogSci*, 2845–2851.

Tian, Y., Ferguson, H., & Breheny, R. (2016). Processing negation without context–why and when we represent the positive argument. *Language, Cognition and Neuroscience*, *31*(5), 683–698.

Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, *4*(1), 7–11.