International Journal of Psychophysiology 110 (2016) 153–162

Contents lists available at ScienceDirect

International Journal of Psychophysiology

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

Neural mechanisms underlying attribution of hostile intention in nonaggressive individuals: An ERP study

Jean Gagnon [a,b,c,*], Mercédès Aubin [a], Fannie Carrier Emond [a], Sophie Derguy [a], Monique Bessette [d], Pierre Jolicoeur [a,c]

[a] Department of Psychology, University of Montreal, Canada
[b] Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal, Canada
[c] Centre de recherche en neuropsychologie et cognition (CERNEC), Canada
[d] Institut Victoria, Montréal, Canada

ARTICLE INFO

Article history:
Received 2 April 2016
Received in revised form 28 June 2016
Accepted 15 August 2016
Available online 16 August 2016

Keywords:
Event-related brain potentials (ERP)
Hostile intention inference
Social cognition
Theory of mind

ABSTRACT

Although the perception of hostile intentions in other people can have a clear adaptive function, researchers have paid little attention to the capacity of nonaggressive individuals to infer hostile intentions in others. The goal of the present study was to study brain mechanisms associated with expectations of hostile/non-hostile intent and their on-line evaluation. Scenarios with a hostile versus non-hostile social context followed by a character's ambiguous aversive behavior were presented to readers, and we recorded and analyzed event-related brain potentials (ERPs) to critical words that disambiguated the hostile versus non-hostile intent behind the behavior. Fifty nonaggressive individuals participated in the study. Non-hostile critical words that violated hostile intention expectations elicited a larger negative-going ERP deflection with central and posterior maximums between 400 and 600 ms after word onset compatible with an N400 effect. Finally, there were marginally significant correlations between N400 effect sizes and hostile as well as neutral attribution bias measured by a self-report questionnaire. The results suggest that nonaggressive individuals evaluate rapidly, on-line, their attributions of the hostile intent of others. The methodology we developed provides the field with a new paradigm with which to study social attributions of hostile intent likely to contribute to hostile or aggressive reactions.

1. Introduction

Imagine the following scenario: You walk into your favorite store and the sales clerk, who is a new employee, says he does not have the item you are looking for. Five minutes later, another customer leaves the store with the item you wanted, after having been served by the same clerk. Your first thought might be that the clerk did not know that he had the item because of lack of experience, and then you might be patient with him. However, if you then learn that he actually does not like you and does not want to serve you, you might feel upset and look for the assistance of another clerk. Now imagine another scenario. Your boss, who likes to take advantage of his employees, decides to give you new responsibilities at the office. You might think that he wants to exploit you again, and be reluctant to accept these new responsibilities. If you then learn that he actually considers promoting you, you will see the situation in a different light and be happy to accept these responsibilities. What these scenarios illustrate is that, when facing ambiguous aversive behaviors from others, we readily infer their intention as being hostile or non-hostile according the social context in which they occur. In social interactions, the capacity to perceive the presence or absence of other people's hostile intentions is essential, not only to initiate/maintain a positive interaction with a benevolent person despite unpleasant behavior, but also to avoid negative interactions or defend ourselves against a malevolent other. The goal of the present study was to study the brain mechanisms that are associated with such expectations of hostile/non-hostile intent and their on-line evaluation. To this end, scenarios with a hostile versus non-hostile social context followed by a character's ambiguous aversive behavior were presented to readers, and we used event-related brain potentials (ERPs) to critical words that disambiguated the hostile versus non-hostile intent behind the ambiguous behavior.

1.1. Attribution of hostile intention in aggressive and nonaggressive individuals

Hostility can be viewed as the cognitive component of aggression since it refers to a wish to hurt or harm others and includes negative attributions to others (e.g. they are stupid; Buss, 1988). In that sense, as

* Corresponding author at: Department of psychology, University of Montreal, P.O. Box 6128, Centre-ville station, Montreal, Quebec H3C 3J7, Canada.
E-mail address: jean.gagnon@umontreal.ca (J. Gagnon).

http://dx.doi.org/10.1016/j.ijpsycho.2016.08.007
0167-8760/

opposed to anger, it does not necessarily involve autonomic reactions or behavioral expression, and must be inferred from the perceiver's perspective. Consequently, the attribution of hostile intention can be viewed as a particular type of inference about someone's mental state and the study of its neural mechanisms can contribute to the theory of mind literature. Social cognitive literature on aggression focuses on two key input variables that influence aggressive behaviors: the features of the person and the features of the situation (Anderson and Bushman, 2002; Crick and Dodge, 1994). Among situational factors, there are aggressive cues, which refer to objects or persons that prime aggression-related concepts in memory, and interpersonal provocations such as insults, verbal or physical aggression, and interference with one's attempts to attain an important goal (Anderson and Bushman, 2002). Among the person factors, there are schemas and scripts, which refer to knowledge structures or interrelated information stored in semantic long term memory (Bushman and Anderson, 2001).

It has been proposed that hostile attributions of intent in aggressive individuals are caused by a chronic activation of knowledge structures representing aggressive schemas and scripts (Crick and Dodge, 1994; Huesmann, 1988). This means that aggressive concepts come to mind spontaneously and exert a more frequent influence on the perception and understanding of the social environment than for nonaggressive individuals (Todorov and Bargh, 2002). Aggressive concepts will be more rapidly activated in the aggressive individual's semantic memory when facing just a few aggressive cues (e.g., ambiguous provocation) and these spontaneous aggressive thoughts will in turn be more likely to generate immediate hostile attribution processes (Anderson and Bushman, 2002). This hypothesis has received partial empirical support from traditional paradigms where hypothetical hostile, ambiguous, or benign social scenarios are presented to subjects. Two studies conducted among undergraduate students and community drivers found that high-aggressive individuals/drivers viewed the intent of the character as more hostile than low-aggressive individuals/drivers in all three situations (Epps and Kendall, 1995) or in ambiguous situations only (Matthews and Norris, 2002); whereas a third study failed to find such a group difference between aggressive individuals and nonaggressive individuals (Helfritz-Sinville and Stanford, 2014). Interestingly, in ambiguous situations, nonaggressive individuals under-attributed hostility and anger to the character even though they perceived the action as intentional (Matthews and Norris, 2002).

Implicit social cognitive paradigms are designed to assess spontaneous dispositional inferences, which occur outside of explicit awareness. For example, in a cued-recall paradigm study, the subjects are instructed merely to read and memorize a series of sentences describing social encounters open to hostile or non-hostile interpretations. By using this paradigm, Zelli and collaborators (Zelli et al., 1995, 1996) found that hostile dispositional cues prompted more recall of sentences than semantic or non-hostile dispositional cues among aggressive subjects, whereas the opposite pattern of results was observed among nonaggressive subjects. These results suggest that nonaggressive individuals do not spontaneously attribute hostile traits to characters while they learn the sentences.

Previous studies leave entirely open the question of how nonaggressive individuals infer hostile intention in others. Current data suggest that, when they have sufficient time, nonaggressive individuals may engage more effortful inferential processes during ambiguous situations involving the search for non-hostile cues while discounting hostile cues. However, at the level of spontaneous appraisal processes, it is reasonable to believe that they make similarly rapid attribution of hostile intention as in aggressive individuals, when the social context contains sufficient evidence for hostile intent (e.g., cues suggesting aggression). In such a context, aggressive-related concepts in nonaggressive individuals' semantic memory would be primed and influence their inferences. An approach that allows a direct measure of intentional processes as they unfold in real time is therefore needed to verify this hypothesis.

1.2. Neural mechanisms underlying intention inferences

The neural processes involved in the ability to understand others' intentions have been studied from two social neuroscience perspectives (see Ciaramidaro et al., 2014). The first perspective proposed that intention understanding is accomplished by means of a motor simulation within the mirror neuron system (Rizzolatti and Sinigaglia, 2010) and used tasks requiring the understanding of intention conveyed by body motion (Becchio et al., 2012). This system includes the premotor cortex (PMC) and the anterior intraparietal sulcus (aIPS; Iacoboni et al., 2005). The second perspective, called theory of mind or mindreading, proposed that understanding intention is related to inferential processes recruited when people reflect on others intentions, without necessary input from the mirror neuron system (Amodio and Frith, 2006). Theory of mind tasks usually involve reading stories or watching cartoons implying beliefs or action goals (Walter et al., 2004). Mindreading about others have been consistently linked to brain network consisting of the medial prefrontal cortex (mPFC), the temporo-parietal junction (TPJ), as well as the adjacent posterior-superior-temporal sulcus (pSTS; Frith and Frith, 2006; Saxe, 2006).

Event-related brain potential (ERP) studies permit a detailed assessment of the time course of the mechanisms involved in social cognition and can inform us about the locus of experimental effects (Bartholow, 2010). As such, ERP techniques represent a valuable approach to verify our general hypothesis that aggressive-related concepts in nonaggressive individuals' semantic memory are temporarily accessible during a hostile social context and generate spontaneous inferences of hostile intent as underlying the ambiguous aversive behavior of others. The N400 ERP component seems relevant to investigate the attribution of hostile intention. The N400 is a negative shift in the ERP waveform that often reaches its maximum amplitude in the neighbourhood of 400 ms after stimulus onset and attains the highest amplitudes at centroparietal electrodes (Kutas and Federmeier, 2011). It can be elicited by presenting a word that is unexpected or fits poorly within a sentence context, reflecting the construction of a situation model by the participant where the information provided by the text is immediately integrated with his conceptual knowledge (Leuthold et al., 2012).

Leuthold et al. (2012) used ERPs to investigate the on-line processing of socio-emotional information. The researchers created prototypical scenarios in which an initial sentence was used to establish a social context, followed by a target sentence containing a critical word that informed the reader of the nature of a character's socio-emotional response to the situation described in the initial sentence. Moreover, the described response either matched or mismatched with what one might have expected given the situation. Results indicated that critical words that mismatched rather than matched with a character's expected feelings elicited a larger ERP negativity, suggesting an N400-like effect (around 200–500 ms after word onset), followed by a larger frontal positivity. This negative-going ERP deflection was stronger over central and posterior, right hemisphere and dorsal electrodes. Dipole source analysis was consistent with sources located in the anterior temporal lobes. Leuthold et al. (2012) interpreted their N400 as reflecting the increased demands of integrating general knowledge about social situations (e.g., scripts) with personal- and context-specific information when there was an apparent mismatch among them. As for the later frontal positivity, this ERP effect was interpreted as the reflection of more effortful processing related, for example, to the suppression of an inappropriately predicted word.

Other studies have demonstrated that an N400 effect may also be found with several other social expectancy violations such as a violation of gender or racial stereotypes (Hehman et al., 2014; White et al., 2009), the prior knowledge of a character's typical behavior (Filik and Leuthold, 2013), an expected emotional outcome (Moreno and