Association for Psychological Science

Perspectives on Psychological Science
2018, Vol. 13(2) 161–165

DOI: 10.1177/1745691618754502
www.psychologicalscience.org/PPS
SAGE

The Error-Related Negativity, Self-Monitoring, and Consciousness

Stanislas Dehaene[1,2]
[1]Experimental Cognitive Psychology, Collège de France, Paris, and [2]INSERM-CEA Cognitive Neuroimaging Unit, NeuroSpin Center, Saclay, France

Abstract
The error-related negativity (ERN) is a negative waveform that arises over the front of the scalp immediately after a participant makes a detectable error. The goal of this short article is to describe my serendipitous encounter with this brain signal in 1993–1994 and to briefly review the operation of the underlying error-monitoring system. Recent work suggests that the ERN reflects an internal comparison, by the anterior cingulate cortex, of two signals: an unconscious representation of the ongoing action and a conscious representation of the intended one.

Keywords
error, metacognition, consciousness, cingulate

Self-monitoring is one of the most remarkable properties of the human brain. For example, when we click on the wrong button or take the wrong turn on the highway, in many cases we immediately realize that we made an error. The *error-related negativity* (ERN) is a brain signal that indexes the brain's self-monitoring circuit. It is a strong negative waveform that arises sharply and immediately after an error, over the front of the scalp, and suggests that the brain immediately noticed its mistake. But this sounds paradoxical: How can the brain make an error and then detect it?

The goal of the present brief article is to review some of the properties of this remarkable error-monitoring system, starting from my early work with Michael Posner and Don Tucker on the "localization of a neural system for error detection and compensation" (Dehaene, Posner, & Tucker, 1994). I was surprised to hear that this article figured among the 30 most-cited articles in APS journals. This was not even a full-length article but merely a technical commentary on a previous article (Gehring, Goss, Coles, Meyer, & Donchin, 1993). It is this original article that deserves the fame, not ours—although we did nail down a few important points about error detection and its relation to the anterior cingulate, which I will explain here.

Ours was a serendipitous finding that occurred during my postdoctoral work in Mike Posner's laboratory at the University of Oregon in 1992. My project was to dissect the sequence of processing stages underlying cognitive operations such as number comparison or word reading (Dehaene, 1995, 1996). To do so, I relied on measurements of the electroencephalogram (EEG), the continuous fluctuations of electrical brain activity on the surface of the head. When we average many EEG signals relative to a fixed time point, such as the presentation of a word, we obtain the *event-related potential* (ERP), a measure of the average time course of brain activity associated with that event. This method provides a simple means of monitoring human brain activity in a noninvasive manner and with millisecond resolution. When I arrived in Oregon, Don Tucker had just designed his by-now widespread "geodesic electrode net," a tight mesh of sponges that allowed for the fast and comfortable application of a large number of electrodes over the entire scalp of adults and even infants (Dehaene-Lambertz & Dehaene, 1994). The nets provide a high spatial resolution, initially with 64, then 128, and ultimately 256 electrodes. We later complemented those recordings with magneto-encephalography (MEG), an advanced physics technique that records

Corresponding Author:
Stanislas Dehaene, Cognitive Neuroimaging Unit, NeuroSpin Center, EA/SAC/DSV/I2BM Bât 145, Point Courrier 156 F-91191 Gif/Yvette, France
E-mail: stanislas.dehaene@cea.fr

magnetic rather than electrical brain signals, with improved sensitivity and spatial resolution. In today's panoply of cognitive neuroscience tools, high-density EEG and MEG recordings play an essential role in analyzing the time course of mental representations (see, e.g., Dehaene, 1996; King & Dehaene, 2014).

Mike Posner and I had long been interested in attention and the mechanisms of self-regulation (Posner & Dehaene, 1994; Posner & Rothbart, 1998). However, we had no plans to study error processing. Our serendipitous finding arose from a keen attention to detail. My first ERP experiment involved a number-comparison task: Participants decided whether a digit was larger or smaller than 5. During the analysis, I carefully pondered each data-processing step (my automatized software pipeline was later incorporated in the first version of Don Tucker's Electrical Geodesics software). In particular, I wondered what to do with error trials, where participants responded "smaller" when they meant "larger," or vice versa. Should they be included or excluded? To inform my decision, I carefully plotted, for each participant, the time course of error trials versus correct trials—and this was my first encounter with the ERN. It was impossible to miss it: Only a few milliseconds after the person pressed the wrong key, there was a huge and sharp negative potential on midline frontal electrodes. The phenomenon seemed to be quite specific to error trials, and it was present in every participant, even those who hardly made any errors—in fact, Gehring et al. (1993) showed that the size of the ERN increases as error probability decreases, a handy trade-off that facilitates its detection.

At that time, I thought that I had made an original discovery, but I was wrong. A few months later, the article by Gehring et al. (1993) appeared, and I learned that we had in fact been scooped by a few years (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1990). Nevertheless, we were the first to obtain high-density recordings of the ERN, and this allowed us to make a number of important points (Dehaene et al., 1994). First, the ERN reflected the operation of an abstract system for error detection, shared by many unrelated tasks. I had data from two very different tasks, number comparison and semantic classification of written words, and after subtracting correct trials from error trials, the ERN was virtually identical in latency and topography in the two cases. It also did not depend on whether the error was committed with the right hand or with the left hand. It seemed that the brain possessed a generic system for error monitoring.

Second, a specific brain area seemed to be responsible for this cognitive function. Gehring et al. (1993), using only five electrodes, could only speculate about its neural sources. With our 64 electrodes, however, we showed that the entire topography of the recordings could be captured by a single source (a single *dipole*, in the jargon of ERPs). Assuming that brain activity can be captured with a few dipoles is obviously simplistic, yet when a single experimental variable is manipulated (here, the difference between error trials and correct trials), I have argued that it may provide a reasonable first approximation (Dehaene, 1996). We concluded that the error effect arose from a specific region called the anterior cingulate cortex (ACC), a region of prefrontal cortex located on the midline of both hemispheres.

Subsequent research has confirmed that ACC is a primary generator of the ERN although probably not the only one (for an in-depth review, see Gehring, Liu, Orr, & Carp, 2012). Using the spatially precise method of functional magnetic resonance imaging (fMRI), Carter et al. (1998) were the first to observe fMRI activity specific to error trials and confined to a small region of the ACC. Later, in a pioneering study using simultaneous EEG and fMRI recordings, Debener et al. (2005) demonstrated a tight correlation between trial-by-trial fluctuations in the size of the ERN and the amplitude of fMRI signals in the ACC. This finding was remarkable: A large set of brain regions distinguished between error trials and correct trials, but the correlation analysis pinpointed ACC as the sole correlate of the early error-detection stage indexed by the ERN.

Back in 1993, why were we so excited about the ACC? One has to remember that this region was essentially discovered through brain imaging. Classical neuropsychology was largely silent about it because of the scarcity of lesions in this region. However, the ACC was systematically observed in a variety of brain-imaging studies that emphasized high-level, effortful, attention-dependent tasks such as the Stroop task (Bush, Luu, & Posner, 2000; Pardo, Pardo, Janer, & Raichle, 1990; Posner, Rothbart, Sheese, & Tang, 2007). By integrating data from brain imaging, neurophysiological recordings, and neuroanatomy, Mike Posner developed the hypothesis that the ACC was involved in executive attention and self-regulation. This was a daring and speculative idea, but the ERN result strongly supported it—this region was obviously involved in self-monitoring. Indeed, a striking aspect of our article, in full agreement with Gehring et al. (1993), was that the ERN seemed to be endogenously computed: The brain labeled its own errors, without requiring any feedback from the experimenter. Furthermore, the ERN emerged so quickly after the key press (or sometimes before it) that it could not reflect any form of sensory or proprioceptive feedback. Somehow, the brain managed to spontaneously detect its own errors, making the ERN a perfect example of endogenous self-regulation.

The hundreds of ERN-related papers that followed have been reviewed elsewhere (Gehring et al., 2012). Here I will comment only briefly on two issues that, 2 decades later, still make the ERN a fascinating brain response: the mechanism by which errors are detected and its relation to consciousness.

A Dual-Route Comparison Mechanism for Error Detection

How is it possible for the same brain to make an error and then immediately "know," without any additional input, that it was wrong? The comparator theory (Coles, Scheffers, & Holroyd, 2001; Gehring et al., 1993) proposed that the ERN arose from a comparison between two internal representations: the actual response and the intended response. Any difference between intention and action signaled an impending error. But the theory faced a problem: If part of the brain knew the correct response, why was the error made in the first place? An ingenious and influential alternative, the conflict-monitoring theory, resolved this conundrum by proposing that the ERN arose solely within the response system from the conflict between several simultaneous response options (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Carter et al., 1998; Yeung, Botvinick, & Cohen, 2004). According to this model, the ERN reflected the amount of energy in the response system: When two responses were simultaneously coactivated, there was a high level of activation, and this signaled an impending error. The ERN was therefore not specific to error trials, but errors were just an extreme on a continuum of motor conflict.

I personally leaned toward the comparator theory because prior training in cognitive neuropsychology gave me ample evidence that the brain often relies on multiple parallel processing routes for the same information (e.g., Coltheart, 1978; McCarthy & Warrington, 1990; Shallice, 1988). However, the conflict-monitoring model seemed more economical and could simulate much, if not all, of the existing data (Yeung et al., 2004). It seemed very difficult to adjudicate between them (for a detailed discussion, see Gehring et al., 2012).

Recently, however, my colleagues and I realized that with EEG and MEG decoding techniques (King & Dehaene, 2014), we could directly test a unique prediction of the comparison model: the existence of two simultaneous routes for the representation of an impending action, one that represented the intention and the other the actual response—and the ERN indexed the comparison of those two representations. Twenty years after my initial work, I therefore returned to my favorite number-comparison task. With my PhD student Lucie Charles, I recorded high-density MEG and EEG signals under speed instructions, such that 10% to 20% of participants' responses were errors (Charles, King, & Dehaene, 2014). We then used machine-learning techniques to decode, from the brain signals, the different types of mental representations that unfolded in the subject's brain. According to the dual-route model, there should be three distinct brain signals, corresponding to three distinct types of mental representations: the actual motor response (left or right hand), the intended response (whether the participant should have pressed left or right), and the quality of the response (correct or error). The conflict-monitoring model, on the other hand, predicted that the correct and incorrect responses were jointly represented within the same neural system and therefore should not be independently decodable.

We trained three independent machine-learning decoders (called *support vector machines*) to decipher, from a segment of EEG-MEG data, each of those three variables on a single-trial basis. Crucially, all three representations could be decoded, with distinct time courses. The most striking finding was the coexistence, in the same brain signals, of distinct representations of the intended and of the actual response. At the very moment when participants hit the wrong key, we could decode from their own brain the response that they should have made. This was direct evidence for the dual-route model. Furthermore, errors could also be decoded (unsurprisingly given the presence of a strong ERN), and, as predicted by the comparison model, the error signal was proportional to the difference between intention and action codes. In particular, the intention code vanished when the digit target was briefly flashed and masked below the visibility threshold. On such invisible trials, participants did not have any conscious knowledge of what they should do (although they still performed better than chance), and as predicted, the ERN also vanished (Charles et al., 2014).

This experiment goes a long way to answer our first question: How does the brain manage to detect its own errors? We propose that it does so by using several parallel decision-making systems (a bit like how the space shuttle sends the same computation to three redundant computers and lets them vote). When the motor system reaches an erroneous decision and clicks the wrong key, the error can be caught by comparing this response with the output of other higher order systems processing the same data (for additional evidence and discussion, see, e.g., Del Cul, Dehaene, Reyes, Bravo, & Slachevsky, 2009; Resulaj, Kiani, Wolpert, & Shadlen, 2009).

Error Detection and Consciousness

A second key issue concerns the relationship of the brain's error response to error awareness: Does it necessarily indicate that the participant is aware of making an error? Or does it tag an unconscious process that detects deviance from the current plan? A beautiful experiment by Nieuwenhuis, Ridderinkhof, Blom, Band, and Kok (2001) provided strong evidence that the error negativity reflects an unconscious process. These authors asked participants to perform a difficult antisaccade task, in which the participants had to move their eyes in the direction opposite a visual target. Unsurprisingly, on some trials, the eyes of the participants briefly moved toward the target (i.e., an erroneous response). Remarkably, in many cases the participants were not aware of their errors—they had no idea that their eyes had transiently been to the wrong place. Even on such unconscious error trials, the ERN continued to be present: The brain detected the error even though the participant did not. Thus, the ERN provides an excellent example of subliminal or unconscious processing. What distinguished conscious from unconscious errors was a later component of the ERP called the *Pe*, standing for *positivity on error*. This is a late and positive brain response, similar to the many components forming a late positive complex, or P3 wave, that frequently appears as a signature of conscious processing (Dehaene, 2014).

Thanks to the Nieuwenhuis et al. (2001) study, the ERN gained renewed fame as a marker of the depth of unconscious processing—it seemed that even prefrontal executive processes could occur without consciousness. Our recent research with Lucie Charles, however, qualifies this idea. In the number comparison task, we masked the target digits such that the same stimulus (say the digit 4) was sometimes reported as subjectively visible and sometimes as invisible (Charles, Van Opstal, Marti, & Dehaene, 2013). The participants had to decide whether the digit was larger or smaller than 5, and because we pressured them to respond very fast, they made a large number of errors—regardless of whether they reported seeing the target. By recording EEG and MEG responses, we then showed that the ERN was present only on conscious trials, not on subliminal ones. When the digits were subjectively invisible, even though the participants clearly processed them, as evidenced by higher-than-chance responding to the number comparison task, their brain no longer detected its own errors.

Although this result superficially seems to contradict the results of Nieuwenhuis et al. (2001), the findings can, in fact, easily be reconciliated (Charles et al., 2014; Charles et al., 2013). We have to carefully distinguish which representation does or does not make it into awareness:

- Consciousness of which action was *performed* is *not* needed for the ERN to arise (Nieuwenhuis et al., 2001).
- Consciousness of which action was *required*, however, *is* needed: There is no ERN when the stimulus is subliminal, such that participants remain ignorant of what they should do (Charles et al., 2013), or in other similar cases where the participants have not yet learned the task or have forgotten it.
- Consciousness of the error itself is *not* indexed by the ERN. Rather, error awareness arises from subsequent processing in a broader network, involving the posterior ACC and many other areas of parietal and prefrontal cortex, and accompanied by a positive ERP (the Pe).

In summary, the ERN is a useful but only partially valid marker of conscious error detection because the error may never make it into awareness. Even on trials where the brain emits an ERN, it may still later conclude (erroneously) that it did not make any error.

In conclusion, the attraction of psychologists and neuroscientists for the ERN can be explained by a combination of factors. First, this is a large brain response, which is easy to detect and to study. Second, the phenomenon is intriguing, almost paradoxical: After all, it is the same brain that makes the error and then detects it. And third, most importantly, it provides a concrete path to attack otherwise difficult issues of self-monitoring and consciousness, which figure among the most important problems that cognitive science must address. I was lucky to figure among its early explorers.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

References

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*, 215–222.

Carter, C. S., Braver, T. S., Barch, D., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*, 747–749.

Charles, L., King, J.- R., & Dehaene, S. (2014). Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. *Journal of Neuroscience*, *34*, 1158–1170. doi:10.1523/JNEUROSCI.2465-13.2014

Charles, L., Van Opstal, F., Marti, S., & Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *NeuroImage*, *73*, 80–94. doi:10.1016/j.neuroimage.2013.01.054

Coles, M. G., Scheffers, M. K., & Holroyd, C. B. (2001). Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biological Psychology*, *56*, 173–189.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151–216). London, England: Academic Press.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience*, *25*, 11730–11737.

Dehaene, S. (1995). Electrophysiological evidence for category-specific word processing in the normal human brain. *NeuroReport*, *6*, 2153–2157.

Dehaene, S. (1996). The organization of brain activations in number comparison: Event-related potentials and the additive-factors methods. *Journal of Cognitive Neuroscience*, *8*, 47–68.

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. New York, NY: Viking Penguin.

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science*, *5*, 303–305.

Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, *370*, 292–295.

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain*, *132*, 2531–2540.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1990). Effects of errors in choice reaction time tasks on the ERP under focused and divided attention. In C. H. M. Brunia, A. W. K. Gaillard, & A. Kok (Eds.), *Psychophysiological brain research* (pp. 192–195). Tilburg, The Netherlands: Tilburg University Press.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*, 385–390.

Gehring, W. J., Liu, Y., Orr, J. M., & Carp, J. (2012). The error-related negativity (ERN). In E. S. Kappenman & S. J. Luck (eds.), *The Oxford handbook of event-related potential components*. doi:10.1093/oxfordhb/9780195374148.013.0120

King, J.- R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*, 203–210. doi:10.1016/j.tics.2014.01.002

McCarthy, R. A., & Warrington, E. K. (1990). *Cognitive neuropsychology: A clinical introduction*. San Diego, CA: Academic Press.

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*, 752–760.

Pardo, J. V., Pardo, P. J., Janer, K. W., & Raichle, M. E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences, USA*, *87*, 256–259.

Posner, M. I., & Dehaene, S. (1994). Attentional networks. *Trends in Neurosciences*, *17*, 75–79.

Posner, M. I., & Rothbart, M. K. (1998). Attention, self-regulation and consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*, 1915–1927.

Posner, M. I., Rothbart, M. K., Sheese, B. E., & Tang, Y. (2007). The anterior cingulate gyrus and the mechanism of self-regulation. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 391–395.

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*, 263–266.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, England: Cambridge University Press.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*, 931–959. doi:10.1037/0033-295X.111.4.939