Behaviour Research and Therapy 62 (2014) 37–46

Contents lists available at ScienceDirect

Behaviour Research and Therapy

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

First steps in using machine learning on fMRI data to predict intrusive memories of traumatic film footage

Ian A. Clark [a, 1], Katherine E. Niehaus [b], Eugene P. Duff [c], Martina C. Di Simplicio [d], Gari D. Clifford [b], Stephen M. Smith [c], Clare E. Mackay [a], Mark W. Woolrich [e], Emily A. Holmes [d, f, *]

[a] University Department of Psychiatry, Warneford Hospital, University of Oxford, United Kingdom
[b] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, United Kingdom
[c] FMRIB Centre, Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford, United Kingdom
[d] Medical Research Council Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom
[e] Oxford Centre for Human Brain Activity (OHBA), Department of Psychiatry, Warneford Hospital, University of Oxford, United Kingdom
[f] Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 31 March 2014
Received in revised form
4 July 2014
Accepted 16 July 2014
Available online 4 August 2014

Keywords:
Intrusive memories
Trauma
Flashback
MVPA
Machine learning
Functional magnetic resonance imaging
Mental imagery

ABSTRACT

After psychological trauma, why do some only some parts of the traumatic event return as intrusive memories while others do not? Intrusive memories are key to cognitive behavioural treatment for post-traumatic stress disorder, and an aetiological understanding is warranted. We present here analyses using multivariate pattern analysis (MVPA) and a machine learning classifier to investigate whether peri-traumatic brain activation was able to *predict* later intrusive memories (i.e. before they had happened). To provide a methodological basis for understanding the context of the current results, we first show how functional magnetic resonance imaging (fMRI) *during* an experimental analogue of trauma (a trauma film) via a prospective event-related design was able to capture an individual's later intrusive memories. Results showed widespread increases in brain activation at encoding when viewing a scene in the scanner that would later return as an intrusive memory in the real world. These fMRI results were replicated in a second study. While traditional mass univariate regression analysis highlighted an association between brain processing and symptomatology, this is not the same as prediction. Using MVPA and a machine learning classifier, it was possible to predict later intrusive memories across participants with 68% accuracy, and within a participant with 97% accuracy; i.e. the classifier could identify out of multiple scenes those that would later return as an intrusive memory. We also report here brain networks key in intrusive memory prediction. MVPA opens the possibility of decoding brain activity to reconstruct idiosyncratic cognitive events with relevance to understanding and predicting mental health symptoms.

Introduction

The focus of the current paper is on using neuroimaging to understand the development of intrusive memories of trauma, that is "recurrent, involuntary and intrusive distressing memories of the traumatic event" (The Diagnostic and Statistical Manual of Mental Disorders, 5th ed.; DSM-5; American Psychiatric Association, 2013). Intrusive memories are a hallmark symptom from the re-experiencing cluster of Post-Traumatic Stress Disorder (PTSD). They have previously been defined as involuntary mental images that occur in a waking state (Frankel, 1994; Jones et al., 2003). Thus, key features of intrusive memories are that they are involuntary rather than deliberately retrieved, i.e. apparently spontaneous (Kvavilashvili, 2014); include perceptual aspects of the traumatic event, i.e. involve mental imagery rather than only verbal thought (Holmes, Grey, & Young, 2005); are in line with episodic and memory recall more broadly (Conway, 2001), and have distressing, i.e. emotional content (Hackmann, Ehlers, Speckens, & Clark, 2004). For example, after a motor vehicle accident, seeing scaffolding smashing through the car windscreen (see Grey & Holmes, 2008; Holmes et al., 2005 for further examples). In their most extreme form, re-experiencing symptoms can present as so-called dissociative 'flashbacks' where patients relive past events as if they are

* Corresponding author.
E-mail address: emily.holmes@mrc-cbu.cam.ac.uk (E.A. Holmes).
[1] Present address: Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London WC1N 3BG, United Kingdom.

http://dx.doi.org/10.1016/j.brat.2014.07.010

happening in the present (American Psychiatric Association, 2013). In contrast, during the experience of an intrusive memory the past events are spontaneously remembered while awareness of the present is maintained.

Due to the nature of this special issue, "*How neuroscience informs behavioural treatment*" within *Behaviour Research and Therapy*, we appreciate that many readers may not have a detailed understanding of neuroimaging terms and techniques. We therefore present a slightly longer than normal introduction to guide the reader through the steps taken before performing the main predictive analysis presented here. We first describe our initial study using traditional neuroimaging analysis techniques (Bourne, Mackay, & Holmes, 2013) and its subsequent replication (Clark, Holmes, Woolrich, & Mackay, submitted for publication). We then introduce the ideas of multivariate pattern analysis (MVPA) and machine learning, before next describing how we utilised these techniques in the current experiment. The aim of this is to provide a methodological basis for understanding the context of the current results and show that these findings are both replicable and reliable. We believe that by using neuroimaging techniques in addition to behavioural, cognitive and psychophysiological experiments we may be able to identify those neural and cognitive functions that are critical for intrusive memory formation. Understanding how intrusive memories are formed from multiple perspectives may allow future work to improve the ability to refine treatments which target the underlying mechanisms of intrusive memory (i.e. symptom) development. Indeed, by gaining the most comprehensive understanding of differences at an individual level, we may be able to open future possibilities of early screening for risk of PTSD, as well as the development of preventative approaches in the immediate aftermath of trauma and for targeted early interventions.

We also note that many different approaches to machine learning and MVPA are evolving, including (but not limited to) Random Forest Theory (Breiman, 2001), Graph theory (Power et al., 2011; Sporns, 2014) and Representational Similarity Analysis (Kriegeskorte, Mur, & Bandettini, 2008), in addition to that used here, a Support Vector Machine classifier (Pereira, Mitchell, & Botvinick, 2009). The current work represents only first steps in applying neuroimaging techniques to understand the neural impact of witnessing trauma and to inform behavioural treatment. We finish by exploring how such techniques might have implications for future cognitive behavioural therapy.

Intrusive memories and PTSD

Most people will experience a traumatic event during the course of their lifetime and a significant minority will go on to develop PTSD (Breslau et al., 1998; Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995). We have successful treatments for the full blown disorder, those recommended by clinical guidelines (e.g. National Institute for Health and Clinical Excellence, 2005) are Cognitive Behavioural Therapy (CBT; e.g. Ehlers & Clark, 2000; Foa & Rothbaum, 1998) and Eye Movement Desensitisation and Reprocessing (EMDR; Shapiro, 1995). However, satisfactory preventative treatments against PTSD development are lacking (Roberts, Kitchiner, Kenardy, & Bisson, 2009). A greater understanding of the brain mechanisms that lead to the development of intrusive memories may help guide effective preventative interventions for the early aftermath of trauma.

We know little, in particular in terms of neuroscience, about why only certain events within a trauma return as intrusive memories when others do not. Processing at the time of trauma (peri-traumatic) is implicated in PTSD development (e.g. Brewin, 2014; Ehlers & Clark, 2000; Ozer, Best, Lipsey, & Weiss, 2003). Additionally, experimental findings implicate heightened emotional processing at the time of the event in intrusive memory development (Clark, Mackay, & Holmes, 2013, 2014). Interestingly, dissociation, defined within the DSM 5 as "a disruption of and/or discontinuity in the normal integration of consciousness, memory, identity, emotion …" (American Psychiatric Association, 2013, p. 291), can be a reaction to extreme emotion, and peri-traumatic dissociation has also been associated with intrusive memory formation (e.g. Daniels et al., 2012; Holmes, Brewin, & Hennessy, 2004). Seminal work on 'flashbulb' memories, defined as 'memories for the circumstances in which one first learned of a very surprising and consequential (or emotionally arousing) event' (Brown & Kulik, 1977) may also illuminate some of the mechanisms involved in intrusive memory formation. While flashbulb memories are a distinct phenomenon (and not exclusive to trauma, but part of autobiographical memory more generally), they may lie on a continuum with intrusive memories. Research suggests that memories that end up as flashbulb memories are psychophysiologically arousing, personally salient and unexpected and sudden (Brown & Kulik, 1977). Indeed, psychophysiology has been associated with intrusive memory development; at the time of viewing a specific film scene that is later recalled as an intrusive memory, heart rate has been shown to drop in comparison to the rest of film viewing (Chou, Marca, Steptoe, & Brewin, 2014; Holmes et al., 2004). Understanding the neural processes involved in intrusive memory formation adds another level of comprehension of this complex phenomenon.

Neuroimaging and established PTSD

The majority of studies using neuroimaging to investigate PTSD have done so once symptoms are already established in patients (Francati, Vermetten, & Bremner, 2007; Hughes & Shin, 2011; Pitman et al., 2012). Neurocircuitry models suggest that PTSD is characterised by reduced activity in the ventromedial prefrontal cortex, which is associated with decision making and emotional response inhibition, and increased activation in the amygdala and other limbic areas, which are associated with emotional processing (e.g. Rauch, Shin, & Phelps, 2006; Rauch, Shin, Whalen, & Pitman, 1998). A further recent model suggests that abnormalities in the amygdala and dorsal anterior cingulate cortex are pre-disposing, while abnormal interactions between the hippocampus and ventromedial prefrontal cortex arise after developing PTSD (Admon, Milad, & Hendler, 2013). While informative for understanding PTSD as a whole, these studies cannot tell us specifically about intrusive memories, that is, those events we need to target within a CBT treatment (e.g. Ehlers & Clark, 2000; Foa, Hembree, & Rothbaum, 2007). Further, studying symptoms once they are already established tells us little about the neural processes involved in intrusive memory *formation* (aetiology).

The trauma film paradigm: an experimental psychopathology approach

Electronic media offers a way to use neuroimaging to investigate the brain responses to experimental analogue trauma exposure and intrusive memory formation. Recent work has examined the effects of electronic media, for example television news film footage, on the development of PTSD symptoms. Individuals exposed for prolonged hours to media footage of terrorist attacks have been shown to present higher scores on stress and trauma related symptom scales both a month after the attack (Holman, Garfin, & Silver, 2014) and 2–3 years after the attack (Silver et al., 2013). Additionally, the DSM 5 (American Psychiatric Association, 2013) now includes exposure to trauma through electronic media in the definition of a traumatic event, with the caveat that the exposure is work related.