Virtual Reality: Treatment Efficacy and a Tool to Study Reactivity in Antisocial Personality Disorder

Ophelia Phillips

D.Clin.Psy. thesis (Volume 1) 2012

University College London

Overview

Part one is a meta-analytic review comparing the efficacy of virtual reality treatments (VRTs) and standard psychological therapies for achieving mental health symptom reduction. Outcomes from twenty-two randomised-controlled trials were quality assessed and meta-analysed. Results indicated that VRTs were equal to, and in some cases superior to comparative treatments, depending on the type of mental health problem being treated. Methodological and heterogeneity issues complicate interpretation. Continued methodologically robust research is required before recommendations for practice can be made with confidence.

Part two presents an empirical study in which virtual reality (VR) was used to investigate emotional reactivity and aggression in antisocial personality disorder (ASPD). Fifteen individuals diagnosed with ASPD and twenty healthy volunteers took part in VR provocation. In response, ASPD participants showed greater negative emotional reactivity, less prosocial behaviour, and a trend towards more aggression than healthy volunteers. Findings tentatively support the notion that ASPD entails difficulties regulating emotions and inhibiting aggression under conditions of perceived threat. Modified large-scale replications are required to substantiate findings. Improved understanding could inform practices for assessing and treating risk of aggression/violence in this population.

Part three is a critical appraisal of the empirical study. It describes my background interest in the research area and critiques a multi-method approach to measurement. The potential for VR to be used as a tool to assess and treat criminogenic risk in ASPD is discussed in more detail. It concludes with personal reflections, highlighting some of the ethical and risk management considerations raised by conducting research with this population.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I dedicate this research to my father, whose big and complicated personality inspired my fascination in human-beings. A huge thank you goes to my incredible husband Tim, who deserves a medal for his relentless love, patience and Thai cooking in the face of my neurotic ways! Thanks also to my dear friends and family for helping me to maintain some semblance of normality and fun throughout this process. A special thanks goes to my baby nephew Adam for bringing bundles of joy during times of monotony.

My sincerest thanks go to Chris Barker, Stephen Butler and Peter Fonagy who supervised this project from the outset. Without Chris's enduring encouragement, dedication, words of wisdom and generous faith in me I would never have survived this process. Stephen's warmth, critical thinking and research conviction helped me to strive for methodological rigour whilst remembering to try and relax. Thanks to Peter who conceptualised the project from the outset and injected his invaluable expertise along the way. I am also hugely indebted to Amanda Williams who encouraged me to go for a meta-analysis and who so generously shared her time and knowledge throughout the literature review process. A big thank you also to Sue Watson, my personal tutor, whose door was always open for a chat and a Kit-Kat!

I am most grateful to the many professionals who made recruitment possible by referring participants and facilitating community leave. In particular I would like to thank Dr Celia Taylor, Dr Cleo Van Velsen, Dr Colin Campbell, Dr Nikki Jeffcote and Professor Anthony Bateman. A huge thanks also to Professor Mel Slater, Dr David Swapp and Aitor Rovira for accommodating me in the virtual reality lab; without their time, flexibility and expertise this project would not have been possible. I am also hugely grateful to Rhianna Jones for her many hours of help with data collection. Finally and most importantly my sincerest thanks go to all the service-users who generously gave up their time to be involved; it just goes to show that stereotypes cannot always be believed!

Part 1: Literature review

Comparative efficacy of virtual-reality treatments and standard therapies for mental health symptomatology: a meta-analysis

ABSTRACT

Background: Research into virtual reality treatments (VRTs) for a range of mental health difficulties has rapidly expanded. There is a need to establish whether the theoretical advantages over traditional therapies translate into symptom reduction.

Objective: To examine the comparative efficacy of VRTs and standard psychological interventions.

Search strategy: Systematic electronic and handsearches from 2001-October 2011.

Selection criteria: Randomised**-**controlled trials, comparing VRTs to established psychological treatments, designed to reduce mental health symptomatology.

Analysis: Studies were synthesised according to (1) type of mental health problem and (2) time point of outcome measurement. Methodological quality was assessed against Cochrane criteria and the Downs and Black checklist (1998). Data was analysed in Review Manager (version 5.0).

Results: Twenty-two RCTs were identified; fourteen treated anxiety disorders, four pain and four body-image distortion/obesity. Eighteen studies (n = 741) provided data for meta-analysis. There were no differences between interventions for anxiety disorder outcomes. Post-treatment, pain symptomatology was significantly reduced following VRT compared with established treatment. Trends were observed favouring VRT for body image distortion/obesity at both post-treatment and follow-up.

Conclusions: In light of heterogeneity and methodological issues, it is tentatively concluded that VRTs are comparable, and in some instances superior to traditional treatments for alleviating mental health symptomatology. Continued efficacy research is needed before recommendations for clinical practice can be made with confidence. Research into theoretical benefits aside from symptom reduction is also required to ascertain the utility and cost-effectiveness of introducing VRT into routine practice.

“Any sufficiently advanced technology is indistinguishable from magic”: Arthur C. Clarke

1. INTRODUCTION

1.1 Virtual reality

Over the past 15 years advances in computer graphics have revolutionised many areas of modern life, ranging from high-tech leisure activities to ground-breaking medical procedures. One important area of advancement is virtual reality (VR), enabling the occupant of a digitalised world to interact with highly realistic environments, blurring the boundaries between the artificial and the real.

VR is “a non-invasive simulation technology that allows a user to interact with a computer-generated environment, in the three dimensions of width, height, and depth” (Rizzo & Kim, 2005). VR is most commonly presented through a head-mounted display containing headphones and screens that provide a first-person perspective. Motion-tracking devices are embedded in the helmet, which monitor the position of the user’s head and adjust the visual imagery, such that the environment responds to their body movements in real-time. An alternative is a computer automated virtual environment, which projects stereoscopic computer-generated images onto three sides of a backlit cubicle. Motion tracking is achieved via an electromagnetic tracking system attached to the user’s shutter glasses. The user can walk freely and naturally through the installation.

The success of a VR application hinges in part on the degree to which the occupant feels ‘present’, as though physically immersed in the environment (Gregg & Tarrier, 2007). This sensation is achieved by shutting out the real world so that only computer-generated stimuli are experienced. The use of visual and auditory stimuli (and to a lesser extent olfactory and tactile cues) add to this sense of reality. Presence can be moderated by individual factors (e.g. the ability to block out

distraction) (Witmer & Singer, 1998) as well as the quality of the VR equipment (Krijn, Emmelkamp, Olafsson & Biemond, 2004).

A number of VR systems are available for commercial use. Costs range from a few hundred to several thousand pounds (plus the cost of a computer). There is scope for purchasers to create and adapt their own environments using 3D software packages (Gregg & Tarrier, 2007).

1.2. VR and psychological research

In the past 15 years, a range of VR applications have been developed for use in psychological and neuropsychological research. VR offers exciting potential to observe and measure brain activity and behaviour during dynamic, complex and realistic situations, whilst ensuring an exacting degree of control over key variables (Bohil, Alicea & Boicca, 2011). Broadly speaking this research has fallen into two domains: (i) studies that aim to understand the processes implicated in typical and atypical brain functioning and behaviour and (ii) studies that investigate the therapeutic application of VR for improving mental health outcomes.

1.2.1 VR process research

Given that the focus of the current review is on VR treatments (VRTs), process studies are not considered in detail. However, it is worth noting that VR-assisted research has contributed to the understanding of dementia (e.g. Flynn, Schaik, Blackman, Hobbs, & Calderdon, 2003), psychosis (e.g. Ku, Cho, Peled et al., 2003) and traumatic brain injury and stroke (e.g. Lee, Ku, Cho et al., 2003). VR has also advanced theories of spatial cognition and navigation (e.g. Astur, St. Germain, Baker et al., 2005), multisensory integration (e.g. Ehrsson, 2007) and social psychology phenomena (e.g. Slater, Antley, Davidson et al., 2006).

1.2.2 VR treatments

In 2002 Norcross, Hedges, and Prochaska conducted a Delphi poll on anticipated psychotherapy trends for the coming decade. A panel of psychotherapy experts ranked VRT third among 38 therapeutic interventions most expected to increase. A plethora of VRTs have since been developed to improve mental health outcomes in psychiatric disorders, chronic and acute pain and neurodegenerative conditions. Presently, these programs are almost exclusively used in research and are yet to filter into routine clinical practice. This is likely to reflect the infancy of VRTs, concerns about cost-effectiveness, clinicians' wariness about heavy reliance on technology, and questions about efficacy. However, VRTs offer some theoretical advantages over standard therapies, including precise control over the degree of exposure to therapeutic scenarios, exposure to scenarios which may otherwise be impractical, and the possibility of tailoring VR environments to meet patients' idiosyncratic needs (Gregg & Tarrier, 2007). For these reasons – coupled with increasingly sophisticated technology and reductions in equipment costs – research into VRTs as a viable treatment option is ever-expanding.

Most VRTs are based on cognitive and/or behavioural models of treatment. However, the therapeutic content varies considerably according to the type of mental health difficulty being treated and the corresponding evidence base. The extent to which a real-life therapist is involved also varies; a few VRTs deliver both cognitive and behavioural elements and may include a 'virtual therapist' meaning that a real-life therapist has minimal involvement. More commonly, VRTs are behavioural and if cognitive components are included in the protocol, they are facilitated by a real-life therapist.

Below the characteristics of three dominant research domains – anxiety disorders, body image disturbance/obesity and pain – are considered. As

aforementioned, VRTs have been developed to treat other types of mental health problems but are beyond the scope of the current review.

1.3 VRTs for specific mental health conditions

1.3.1 Anxiety disorders

VRTs have been most extensively researched in the field of anxiety disorders (Powers & Emmelkamp, 2008). These programs are invariably VR exposure treatments (VRET's) in which the user is gradually exposed to computer-generated internal (e.g. interoceptive arousal) and external (e.g. a phobic object) feared stimuli. Some treatment packages include cognitive components that are delivered by a real-life therapist during VRET and/or in concurrent sessions. VRTs have been developed for post-traumatic stress disorder (e.g. Difede & Hoffman, 2002), panic disorder (Choi, Vincelli, Riva, Wiederhold, Lee, & Park, 2005) social phobia (e.g. Klinger, Bouchard, Légeron, Roy, Lauer, Chemin, & Nugues, 2005) and a range of specific phobias including aviophobia (e.g. Rothbaum, Hodges, Smith, Lee, & Price, 2000), acrophobia (e.g. Emmelkamp, Bruynzeel, Drost & van der Mast, 2001) and arachnophobia (e.g. Carlin, Hoffman & Weghorst, 1997).

The therapeutic content reflects the fact that exposure treatment for anxiety disorders shows some of the largest effect sizes in the literature (Deacon & Abramowitz, 2004) and VRET is seen as a natural extension of this (Emmelkamp et al., 2001) with some additional benefits (e.g. Côté & Bouchard, 2008). For some anxiety disorders, in vivo exposure is impractical and potentially dangerous (e.g. driving phobia) whilst for others the cost is restrictive (e.g. aviophobia). Patients may also perceive VRET to be less aversive because they are aware that the technology can be switched off (Gregg & Tarrier, 2007). VRET can also recreate situations that cannot be re-experienced in vivo such as combat situations (e.g. Rothbaum et al., 2001). This may be useful for patients with PTSD whose treatment is otherwise dependent on their ability to visualise during imaginal exposure. Since avoidance is

a core feature of many anxiety disorders, VRET may also prove more effective than imaginal exposure as it provides less opportunity for covert avoidance (Tarrier et al., 1999).

1.3.2 Pain

There is a growing body of research into VRTs for pain reduction (Botella et al., 2008). These are distraction-based programs, in which the patient enters a pleasurable virtual world and is instructed to engage in a cognitive task. For example, Hoffman et al. (2006) developed 'SnowWorld' in which the user glides through an icy canyon and pushes a keyboard button to shoot virtual snowmen. VR distraction protocols are predominantly behavioural and do not necessitate a therapist's involvement. They have been developed to reduce experiences of acute and chronic pain across a range of contexts, including induced experimental pain (e.g. Hoffman et al., 2006), procedural pain (e.g. Nilsson, Finnstrom, Kokinsky, & Enskar, 2009) and chronic health conditions (e.g. Leibovici, Magora, Cohen, & Ingber, 2009).

The theoretical rationale is that distraction possesses considerable efficacy in pain reduction (Blount, Piira & Cohen, 2003) and it is one of the most commonly used psychological methods for analgesia (Botella et al., 2008). Treatment efficacy is thought to reflect the fact that because humans have finite attentional resources, a distraction task that consumes some portion of those resources leaves less cognitive capacity available for processing pain (McCaul & Malott, 1984).

VR may lend itself to analgesia as it ideally lures attention into the computer-generated world, leaving less attention available to process incoming nociceptive signals (Hoffman et al., 2004). Moreover, VR possesses the characteristics of a good distractor, appealing to multiple sensory modalities and provoking emotional involvement of the individual (Wismeijer & Vingerhoets, 2005). VR distraction also shares the theoretical advantages associated with VRET, including the potential to tailor scenarios to meet patients' idiosyncratic preferences, and to develop scenarios

that would otherwise be unaffordable/impractical (e.g. visiting snow canyons). This may be particularly important for patients with chronic health conditions, where associated disabilities would preclude these activities in reality.

1.3.3 Body-image distortion/obesity

A less well researched but emerging application of VRT is in the field of obesity, with the aim of improving body-image and facilitating weight loss. The focus on body-image emerges from the observation that obese populations frequently suffer from body-image distortion (Friedman & Brownell, 1995), which has repercussions for quality of life and weight gain (Riva, Bacchetta, Baruffi, & Molinari, 2001).

The therapeutic content of VRTs in this domain includes: (i) relaxation in which the user is immersed in a therapeutic scenario (e.g. a green valley) and, (ii) temptation and exposure to triggers of overeating episodes (e.g. a full refrigerator). The rationale behind relaxation is that stress and negative emotionality are considered critical factors in overeating and preference for calorific foods (Geliebter & Aversa, 2003), representing a maladaptive coping strategy (Crowther, Sanftner, Bonifazi, & Shepherd, 2001). Traditional obesity treatments therefore include CBT-based relaxation for managing stress and shifting emotions (Ong, Linden & Young, 2004). The rationale behind exposure to overeating triggers emerges from CBT models of addiction (e.g. Beck, Wright, Newman & Liese, 1993) in which internal and external high risk situations trigger cravings, undermine control and provoke relapse (Marlatt & Gordon, 1985). Accordingly, CBT packages for obesity (e.g. Cooper Fairburn & Hawker, 2003) facilitate relapse prevention by including exposure and response prevention.

Another component of VRT for body-image distortion relates to the VR equipment itself, which creates alterations in the occupant's sensorimotor loops (Riva & Melis, 1997). These alterations are unintentionally triggered in almost all VR

systems due to distortions, time delays and noise. The somesthetic system contains a proprioceptive subsystem that senses the body's internal state, including the position of limbs/joints and muscle tension (Sadowsky & Massof, 1994). Mismatches between the signals from the proprioceptive system and the external signals of VR lead to greater awareness of the perceptual and sensory processes associated with the way the body is experienced. When new information that is incompatible with existing body schema is introduced during VR, there may be scope to influence the user's body-image representations.

When treating body-image distortion/obesity, VR relaxation may be preferable to traditional relaxation, as it is not dependent on an individual's capacity to produce relaxing images (Vincelli, 1999). Furthermore, capitalising on the inherent psychophysiological effects of VR on body schema may offer an effective means of treating body-image distortion, which is conceptualised as a complex condition (Thompson, Heinberg, Altabe, & Tantleff-Dunn, 1999).

1.4 Previous reviews of VRT efficacy

A central question when establishing a new treatment is the degree to which it achieves intended outcomes. Outcomes vary from treatment viability, to uptake rates, to effects on global functioning. Perhaps most importantly, treatment efficacy for achieving symptom reduction requires careful and systematic evaluation. Over the past decade, VRT efficacy and effectiveness research for symptomatology outcomes has rapidly expanded and been synthesised into a number of narrative and meta-analytic reviews (see Table 1).

1.4.1 Anxiety disorders

Early qualitative reviews (e.g. Pull, 2005; Krijn, Emmelkamp, Olafsson, Biemond, 2004) summarised the results of case studies, open clinical trials and uncontrolled designs and concluded that VRET showed promise in treating symptomatology in a range of specific phobias (i.e. acrophobia, claustrophobia,

aviophobia, and arachnophobia). More recent reviews (e.g. Powers & Emmelkamp, 2008; Parsons & Rizzo, 2008; Price, Anderson & Rothbaum, 2008; Meyerbroker & Emmelkamp, 2010) have synthesised findings from controlled trials. Consistent with earlier speculations, VRET out-performed no-treatment conditions in aviophobia, acrophobia and arachnophobia, both in primary symptomatology and secondary domains (Powers & Emmelkamp, 2008). In terms of comparative treatment effectiveness, narrative reviews indicate that VRET is at least as effective as traditional exposure for a range of specific phobias (Cote & Bouchard, 2008; Meyerbroker & Emmelkamp, 2010). Two meta-analyses have also obtained small but significant effects favouring VRET over established treatments for phobias (Powers & Emmelkamp, 2008; Parsons & Rizzo, 2008).

For more complex anxiety disorders, the evidence favouring VRET over no-treatment conditions is also encouraging. Early narrative reviews (e.g. Pull, 2005) summarised findings from several studies of PTSD and social anxiety disorder and concluded that there was preliminary evidence for VRET effectiveness. More recently, two meta-analyses (Parsons & Rizzo, 2008; Powers & Emmelkamp, 2008) found that VRET reduced symptoms of PTSD, social phobia and panic disorder compared to no treatment conditions. Research into comparative efficacy for these more complex anxiety disorders is less consistent. Although Parsons and Rizzo (2008) and Powers and Emmelkamp (2008) found significant effects favouring VRET over in vivo exposure for PTSD, panic disorder and social phobia, findings were limited by the small combined sample sizes and high proportion of non-randomised trials included in their meta-analyses.

Meyerbroker and Emmelkamp (2010) conducted the most recent review of the VRET anxiety disorders literature, systematically evaluating the evidence from RCTs. Consistent with previous findings, VRET was found to be efficacious for specific phobias, comparable – but not necessarily superior – to traditional CBT. The

review provided a more sobering look at VRET efficacy for other anxiety disorders, noting the lack of comparative treatment trials for PTSD and other complex anxiety disorders.

1.4.2 Pain

Botella et al. (2008) reviewed case-studies, clinical trials and descriptive studies for VR distraction in the field of pain management. They concluded that immersive VR distraction is a promising technique for treating acute procedural pain by way of reducing perceptions of pain, unpleasantness and distress. However, the authors cautioned against definitive conclusions, highlighting an absence of controlled trials with large samples, and the paucity of research into chronic pain management. Shahrbanian, Ma, Korner-Bitensky and Simmonds (2009) narratively reviewed 27 RCTs and descriptive studies of VRT. Consistent with Botella et al. (2008), they found strong evidence to support immersive VR distraction over no-treatment conditions for acute pain management, but noted an absence of empirical research regarding chronic pain and a paucity of comparative treatment trials.

Malloy and Milling (2010) described evidence from 11 RCTs and mixed-model studies that compared VR distraction to control conditions and alternative interventions. They concluded that there was solid evidence for the efficacy of VR distraction in treating experimental and burn injury pain. However, only three comparative psychological treatment trials were reviewed, each with relatively small sample sizes. Additionally there was a lack of available evidence for procedural and chronic pain, limiting conclusions about efficacy in these domains.

1.4.3 Body-image distortion/obesity

Riva and Molinari (2009) narratively reviewed case studies, uncontrolled trials and RCTs of VRT effectiveness for eating disorders. They concluded that VR appeared to be an effective means of reproducing everyday situations that provoke negative emotional responses in relation to food. They also concluded that VRT

outperformed traditional CBT and psycho-nutrition for reducing body-image distortion and improving self-efficacy. However, the review did not critically appraise included studies, nor take into account methodological limitations when drawing conclusions about effectiveness.

Gregg and Tarrier (2007) included outcome data from body-image distortion/obesity studies in their trans-disorder review of VRT. They tentatively concluded that VRT was associated with greater body-image satisfaction and self-efficacy in the short- and long-term, compared with control conditions and CBT. However, no differences were observed between active treatments in the amount of weight loss achieved. Gregg and Tarrier (2007) highlighted the need for more controlled trials in order to draw conclusions about VRT efficacy in this domain.

1.5 The current review

1.5.1 Aims and rationale

This review provides a systematic overview of existing research into the comparative efficacy of VRTs and established psychological treatments for achieving symptom reduction across mental health conditions/problems. Primary symptom reduction was selected as the outcome of interest as this is a key objective in treatment. Other important outcomes (e.g. treatment satisfaction) require evaluation, but are beyond the scope of this review.

Given that several previous reviews have concluded that VRTs out-perform no-treatment conditions (e.g. Gregg & Tarrier, 2007; Powers & Emmelkamp, 2008), reviewing comparative efficacy is the next logical step. Although Cote and Bouchard (2008) point out that VRTs were developed to address the limitations of traditional treatments rather than outperform them per se, there is a need to establish whether these theoretical benefits translate at a symptom reduction level. This could inform

decision-making about the utility and cost-effectiveness of implementing VRTs in routine clinical practice.

Although a number of previous reviews (see Table 1) have synthesised VRT efficacy data, there is a need for a fresh look at the existing evidence for a number of reasons. First, with the exception of Meyerbroker and Emmelkamp (2010), previous reviews have included studies that fail to meet stringent methodological criteria, thereby limiting empirical decision-making. Second, most of the earlier studies reviewed contained small sample sizes and made inadequate use of null hypothesis significance testing (Parsons & Rizzo, 2008). Conversely, studies in the present review are limited to RCTs that include samples of 10 or more per condition. Third, most of the RCTs have historically been limited to specific phobias; recent research regarding the efficacy of VRTs for treating other disorders/conditions requires evaluation. Finally, the majority of previous reviews did not quality assess studies or meta-analyse outcomes.

Since there are still relatively few large-scale studies on the effects of VRT on mental health symptom reduction, a meta-analytic review may be the most appropriate method for synthesising data. Such analyses provide estimates of a population effect size across independent studies. They increase statistical power to detect true nonzero population effects by lowering the standard error, thereby narrowing the confidence intervals associated with the population effect size estimate (Cohn & Becker, 2003). Hence, a meta-analysis – as opposed to a qualitative review – might facilitate a better understanding of the comparative efficacy of VRTs and established psychological treatments.

The conceptual focus of the present review differs from previous reviews in that it includes studies generated from across mental health problem domains that applied a range of VRTs. Two previous reviews (Glantz, Rozzo & Graap, 2003; Gregg & Tarrier, 2007) synthesised trans-disorder data but the body of efficacy research has significantly expanded since their publication.

The present review assesses the quality of published RCTs using Cochrane criteria and the Downs and Black (1998) checklist. In light of these findings meta-analytic methods are applied to ask:

- Are there differences between VRTs and standard psychological treatments on symptomatology outcomes across mental health conditions/problems, both at post-treatment and long-term follow-up?

Table 1: Reviews into the effectiveness of VRTs for improving mental health outcomes conducted between 2001 and October 2011

Author (date)	Population/problem	Research question	Method	Main Findings	Difference from current review
ANXIETY DISORDERS					
Bush (2008)	Anxiety disorders	The viability of VRET for treating anxiety disorders.	Narrative	VRET is effective for treating specific phobias compared to control conditions. Shows promise in other anxiety disorders, but there is limited research and absence of RCTs. Need to reduce the costs of VRE in order to increase uptake.	Limited to anxiety disorders and exposure. Included case studies and small *n* designs. Did not evaluate comparative treatment efficacy.
Coelho et al. (2009)	Acrophobia (fear of heights)	The utility of VRET in the research and treatment of acrophobia.	Narrative	VRET is well established effective treatment. It offers theoretical benefits over in vivo exposure, including better control over the situation, avoiding public embarrassment and preservation of confidentiality. VR has also contributed to the understanding of acrophobia: e.g. it is motion combined with height, rather than height per se, that triggers phobic response.	Limited to acrophobia. Reviewed process studies as well as treatment studies. Reviewed case studies in addition to controlled trials.
Côté & Bouchard (2008)	Specific phobias	Effectiveness and efficacy of VRET compared to control conditions and traditional treatments	Narrative	VRET outperforms control conditions and appears to be at least as efficacious as traditional exposure. However, more studies with stronger methodological criteria are needed to	Limited to phobias. Included non RCTs and case studies. Primarily reviewed VRET in comparison to control groups rather than active treatment

				fully form conclusions about comparative treatment efficacy.	conditions.
Da Costa et al. (2008)	Aviophobia (fear of flying)	The effectiveness of VRET for treating aviophobia.	Narrative	VRET appears to be effective in aviophobia treatment and out-performs control conditions. VRET effective with or without CBT and/or psychoeducation. There is a need for more RCTs to assess efficacy.	Only focussed on aviophobia. Included non-randomised trials and no-treatment control groups. Was not seeking to evaluate comparative treatment efficacy.
Da Costa et al. (2010)	Driving phobia	The effectiveness of VRET for treating driving phobia.	Narrative	VRET shows promise in reducing anxiety associated with driving phobia. It is cost-effective and reduces some of the risks associated with in vivo exposure. However, for some patients VRET may not be sufficient in isolation. More RCTs are required to fully evaluate efficacy.	Only focused on driving phobia. Included uncontrolled trials (case studies) and no-treatment conditions. Was not seeking to compare VRET to other psychological treatments.
Krijn et al. (2004)	Anxiety disorders	Effectiveness of VRET for treating anxiety disorders.	Narrative	Evidence for the effectiveness of VRET for treating aviophobia, acrophobia and arachnophobia. Requires more RCTs with larger samples, comparing VRET to standard exposure. Also, need trials that assess VRET as standalone treatment rather than in conjunction with other treatments.	Limited to anxiety disorders. Included case studies. Did not directly set out to assess comparative treatment efficacy.

Meyerbroker & Emmelkamp (2010)	Anxiety disorders	The efficacy of VRET for treating anxiety disorders (controlled studies only). Also investigated process/mechanisms of change underlying VRET.	Systematic	There is only strong evidence for the effectiveness of VRET compared to established treatments in treating specific phobias. Results are also promising for more complex anxiety disorders, but controlled trials are required to inform empirical decisions. Many studies used VRT in conjunction with other treatment techniques; thus more research is required to dismantle the effective ingredients.	Only anxiety disorders and VRET. Included no-treatment conditions in addition to comparative treatments.
Parsons & Rizzo (2008)	Anxiety disorders	The effectiveness of VRET on affective outcomes for treating anxiety disorders (PTSD, social phobia, acrophobia, aviophobia, arachnophobia)	Meta-analysis	Large decline in affective anxiety symptoms across all disorders following VRET compared to control groups and comparative treatments. However, some sample sizes small (e.g. PTSD) and there was an absence of information available for moderator analysis. More research needed for uniform and detailed understanding of moderators to treatment effectiveness (e.g. presence, immersion, duration of anxiety disorder).	Only anxiety disorders and VRET. Only focused on affective outcomes, not necessarily primary symptom reduction. Included case reports in addition to RCTs.
Powers & Emmelkamp (2008)	Anxiety disorders	The efficacy of VRET compared to in vivo exposure and control conditions (e.g. wait-list) for treating primary and	Meta-analysis	Large effect size for VRET compared to control conditions in primary symptom reduction and secondary domains (e.g. general functioning). Also small but significant effect	Limited to anxiety disorders and VRET. Included studies that compared VRET to no-treatment/wait-list control in addition to traditional

		secondary outcomes in anxiety disorders (specific phobias, social phobia, panic disorder, PTSD).		favouring VRET over in vivo exposure. A dose-response effect was found; more VRET sessions were associated with larger effects.	treatments.
Price et al. (2008)	Aviophobia	The effectiveness of VRET for treating aviophobia.	Narrative	VRET effective in treating aviophobia. Comparable to established treatments, i.e. In vivo and cognitive therapy. However, studies that strictly used VRET protocol obtained weaker effect sizes than those that incorporated cognitive elements. VRET offers some benefits (e.g. privacy) over traditional treatment.	Only focused on aviophobia and VRET. Included case studies in addition to controlled trials. Included no-treatment comparison groups.
Pull (2005)	Anxiety disorders	The effectiveness of VRET for treating anxiety disorders.	Narrative	VRET appears to have potential for treating a range of anxiety disorders. However, there is an absence of controlled trials; further research is required to confirm VRET efficacy.	Only anxiety disorders and VRET. Included small *n* designs, uncontrolled trials and case studies. Not seeking to directly assess comparative treatment efficacy.
PAIN					
Botella et al. (2008)	Pain	The effectiveness of VR distraction for reducing pain during medical procedures.	Narrative	VR distraction is a promising technique for pain control in medical procedures. However, there is a need for more controlled trials with larger samples, using subjective as well as objective measures of pain.	Focused exclusively on VRTs for pain. Included case studies and un-controlled studies. Was not trying to evaluate efficacy in relation to other types of psychological treatment.

Mahrer & Gold (2009)	Pain	The effectiveness of VRT for pain control	Narrative	VRT for pain is in infancy but shows promise in specific conditions and acute procedural pain. More research required with sound methodology and large samples to draw firm conclusions on effectiveness.	Focus on pain studies only. Reviewed case studies and non-controlled trials in addition to RCTs.
Malloy & Milling (2010)	Pain	Effectiveness of VR distraction in pain reduction in controlled trials only.	Narrative	VR distraction effective at reducing experimental pain and burn injury pain compared to non-VR distraction and NT. Inconsistent findings for needle-related pain. More sophisticated VR technology is associated with greater pain reduction.	Only pain studies reviewed. Included no-treatment control groups in addition to non-VR distraction and hypnosis.
Shahrbanian et al. (2009)	Pain	The effectiveness of VR for pain management (RCT and descriptive studies only).	Systematic	There is strong evidence supporting immersive VR for treating acute pain but limited evidence for reductions in chronic pain. More RCTs and more research into non-immersive VR approaches are required.	Pain studies only. Included descriptive studies. Review did not directly seek to evaluate comparative treatment efficacy.
Body image/ obesity					
Riva & Molinari (2009)	Eating disorders/ obesity	The effectiveness of VRET for reducing negative emotions related to food and body-image.	Narrative	VRET appears to be a valid means of simulating everyday situations that provoke negative emotional responses in patients with eating disorders/obesity. It also appears to facilitate improvements in body-image	Eating disorders only. Included case studies and uncontrolled trials in addition to RCTs. No quality assessment or meta-analysis performed.

				and self-efficacy over and above traditional treatments.	
Trans-disorder					
Glantz et al. (2003)	Mental health problems (various)	The progression of VRT for mental health problems.	Narrative	VRTs have been developed for wide range of mental health problems. There is evidence of effectiveness in treating specific phobias. Some promising evidence for treating PTSD but mostly from case studies. VRT offers potential for treating eating disorders and pain but research is in infancy. The lack of comparative treatment trials in pain precludes conclusions about cost-effectiveness. VRTs for social phobia are only just emerging.	Many uncontrolled trials and case studies. Lack of comparative treatment studies across all mental health domains. No meta-analysis performed.
Gregg & Tarrier (2007)	Mental pealth problems (various)	A review of the current status of VRT in the mental health arena.	Meta-analysis	VRTs are more effective than no-treatment conditions but the available data does not support the effectiveness of VRT over traditional forms of treatment. There is a need for controlled trials involving clinical populations in order to assess the comparative efficacy of VRTs.	Many studies reviewed were case studies. Was seeking to assess the efficacy of VRT compared to no-treatment conditions in addition to other treatments.

2. METHOD

2.1 Inclusion and exclusion criteria for studies in this review

Studies were selected according to the following apriori inclusion criteria:

- Scope of studies:
  1. Published between 2001 and October 2011
  2. English language
  3. Peer reviewed journal articles

- Design of studies:
  1. Randomised-controlled trials, comparing VRT efficacy with at least one other active psychological treatment
  2. A sample size of >10 in each active treatment group

- Types of participants:
  1. Adult, adolescent and child populations;
  2. Suffering from a mental health condition/psychological problem requiring intervention
  3. Recruited from clinical (e.g. hospital) and/or non-clinical (e.g. students) settings

- Types of interventions:
  1. One or more session(s) of any type of VRT (e.g. VRET, VR distraction,) and;
  2. One or more session(s) of established psychological treatment (e.g. CBT) routinely delivered for the presenting problem under investigation
  3. Studies that had a non-active arm (e.g. wait-list) as an adjunct to active treatments were also included

- Types of primary outcome measures:
  1. At least one validated measure or visual analogue scale of symptom severity relevant to the presenting problem(s)
- Excluded studies:
  1. Reviews/meta-analyses, or other non-primary research
  2. Studies in which the comparative treatments were non-psychological (e.g. pharmacological treatment in isolation)

2.2 Search Strategy

Four procedures were used to identify all relevant studies published between 2001 and the cut-off date of October 2011.

1. The PsychInfo and Medline databases were searched within the specified timeframe.
2. Previous literature reviews in the field of VR and mental health were consulted. These reviews were identified during electronic searches and by consulting The Cochrane and Campbell Collaboration databases and hand-searching Clinical Psychology Review in the last two years.
3. Within these databases, two key journals (i.e. those that produced the most hits relevant to the subject area) were hand-searched over the past two years: Annual Review of Cybertherapy & Telemedicine, and Cyberpsychology & Behaviour.
4. The reference lists of identified articles were scanned to detect any studies that might have been missed during the electronic search.

2.2.1 Electronic search of databases

A systematic search of the PsychInfo and Medline databases was carried out, limiting results to peer reviewed journals (applied post-hoc in Medline) and English language articles published between 2001 and October 2011.

Initial searches included broad search terms to identify studies that investigated the application of VR in the mental health arena. Terms such as "computer-assisted treatment" were not included in the final search strategy as they yielded an over-inclusive group of studies (e.g. computerised CBT) and failed to elucidate any relevant studies over-and-above more stringent search terms (e.g. virtual reality). Additional problem-specific terms such as anger/psychosis/schizophre*/developmental disorders (the asterisk is a wildcard convention used to encompass variant terms such as schizophrenia, schizophrenic etc.) were initially included but failed to produce any relevant hits and were therefore dropped.

To maximise specificity and relevance of hits, the final search terms (see Table 2) delineated (a) clinical condition/problem, (b) intervention type, (c) comparison group and (d) outcome of interest (efficacy as opposed to process-driven studies). Terms were entered as individual 'Keyword' searches and thereafter in combination. Individual search terms were also exploded and combined (e.g. mental health *and* virtual reality) in the PsychInfo and Medline thesauruses but yielded no new relevant studies. This indicated that the chosen search terms sufficiently captured all relevant studies in the field.

Table 2: Electronic search terms

Search term category	Terms applied
Condition/problem	anx*/depress*/phobi*/psychol* disorder/mental health/ post traumatic/ pain/eating disorder/obes*
Intervention	virtual reality/virtual environment
Comparison	compar*/control group/cognitive behavio?r /in vivo exposure/treatment as usual
Outcome	intervention/treat*/therap*/efficacy/effectiveness/symptom*

2.3 Data collection and analysis

2.3.1 Screening and selection

All studies were first screened for content relevance by title and were included if they referred to any VRT for mental health conditions/problems. Thereafter detailed checks were made against selection criteria based on the abstracts and full text of articles.

2.3.2 Data extraction

Data were extracted from full text copies of studies that met inclusion criteria and organised in a data extraction form (see Table 3) according to the problem/condition being treated. Information extracted included sample size, age, gender mix, length of treatment, number of sessions, VRT treatment components, comparative treatment components, diagnostic and outcome measures, length of follow-up and the quality rating obtained on the Downs & Black checklist (1998).

2.3.3 Assessment of risk of bias

Risk of bias was assessed in two ways. Firstly according to the standard approach described in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins & Green, 2005), which considers sequence generation,

allocation concealment, blinding of assessors and reporting of loss to follow-up. Secondly, methodological quality was critically appraised using the Downs and Black checklist (1998) (see Appendix 1). This checklist was developed to evaluate the quality of randomised and non-randomised studies of healthcare interventions. It includes 27 items divided into 5 subscales: reporting (10 items), external validity (3 items), internal validity bias (7 items), internal validity confounding (6 items), and power (1 item). A maximum score of 32 is obtained, with higher scores representing better methodology.

The Downs and Black checklist (1998) possesses good test-retest reliability (r = .88) and internal consistency (Kuder-Richardson formula 20 = .89). The item concerning blinding of participants to treatment condition is not relevant to psychological intervention research, as it is rarely possible to conceal from participants which treatment they are receiving (Roberts, Kitchiner, Kenardy & Bisson, 2010). Additionally, due to the inherent characteristics of RCTs, scores on the external validity items are typically low.

2.3.4 Data synthesis

Studies were classified according to (1) the type of problem/condition being treated and (2) time-point of outcome measurement, distinguishing between those that utilised a pre-post design and those that included long-term follow-up.

2.3.5 Measures of treatment effect

Review Manager 5.0 was used to calculate overall estimates of treatment effect with 95% confidence intervals (negative estimates representing results favouring VRTs). Owing to the heterogeneity of outcome measures (utilising different scales of measurement) continuous data were analysed using standardised mean difference (SMD). Separate post-treatment comparisons were run for the three categories of mental health problems that emerged from the literature. Where

applicable, separate comparisons were also conducted for follow-up data, in order to investigate comparative longitudinal treatment effects.

2.3.6 Unit of analysis issues

Meta-analytic methods were applied to contribute to the review question; namely to establish the comparative efficacy of VRTs versus established psychological treatment at achieving primary symptom reduction in the short and longer term. Consequently, data from other outcome domains (e.g. treatment satisfaction) were not included in the analysis.

Where studies used multiple outcome measures of symptomatology, the meta-analysis included the main measure if this was reported by the authors. Where this information was not readily available, the most widely used and/or validated measure was selected for analysis.

In three-armed (or more) trials that included wait-list/no-treatment conditions, only active treatments were entered into the meta-analysis. The rationale behind this is that previous reviews of the literature (e.g. Powers & Emmelkamp, 2008) have provided evidence favouring VR-based treatments over no treatment conditions; therefore it was not considered worthwhile repeating this analysis. In three-armed trials that included more than two active treatments, the most well-established comparative psychological treatment (and best fit to its VRT counterpart) was selected for inclusion in the analyses.

2.3.7 Assessment of heterogeneity

Forest plots were visually inspected to explore for heterogeneity. Heterogeneity between studies was also measured by observing the I squared test and the chi-squared test ($p < 0.10$). An I-squared of less than 30% was considered to indicate mild heterogeneity and in these instances a fixed-effects model was

applied to synthesise the results. This model makes the assumption that there is one single average effect (the comparative efficacy VRT) and that studies which are combined come from a population measuring this fixed effect. An I-squared of 30% - 50% was taken to indicate moderate heterogeneity. An I-squared of above 50% was considered to indicate notable heterogeneity and in these instances a random-effects model was used to synthesise results.

2.3.8 Assessment of publication bias

It was decided apriori that if sufficient studies were available, funnel plots (of treatment effects estimated against the sample size of the studies) would be prepared and examined for signs of asymmetry. Where asymmetry was identified, possible reasons for this would be considered.

3. RESULTS

3.1 Results of the search

The combined electronic searches produced 136 references. Studies were immediately excluded if they were not RCTs (e.g. Penate, Pitti, Bethencourt, Fuente, & Gracia, 2007), focused exclusively on mediators (e.g. Krijn, Emmelkamp, Biemond et al., 2004) and/or moderators (e.g. Gorini, Mosso, Mosso, Pineda et al., 2009), or concerned outcomes other than symptomatology (e.g. St-Jacques, Bouchard & Belanger, 2010). Thereafter the abstracts of 64 articles identified for potential inclusion were consulted. This left 39 references, which were read in entirety for more detailed checks against inclusion/exclusion criteria. This procedure identified 22 eligible studies to be included in the review. Sixteen of these were identified from electronic databases and six (Leibovici et al., 2009; Mühlberger et al., 2003; Patterson et al., 2006; Riva et al., 2001; Riva et al., 2002; Wiederhold et al., 2002) were sourced elsewhere (e.g. reference lists). The main reasons for excluding the 17 trials were insufficient sample sizes and non-randomised allocation to treatment conditions. A flow diagram is presented in Figure 1.

3.2 Description of included studies

3.2.1 Design of studies

By virtue of the selection criteria, all 22 studies were RCTs with two or more active treatment arms. None of the participants were blinded to their treatment group. Sample sizes of the included studies varied between 20 (McLay et al., 2011) and 211 participants (Riva et al., 2006), totalling 1192. Full details of included studies are described in Table 3.

3.2.2 Scope of studies

Of the 22 studies included in the review, 13 were conducted in European countries, five in the United States, two in Canada, one in Asia and one in Australia. Fourteen studies recruited participants with anxiety disorders, four with body-image distortion/obesity and four with pain. The majority recruited from hospital/clinic settings (*n* = 12) and the remainder from community (*n* =7) and student populations (*n* = 3). All studies recruited adult samples except for Gershon et al. (2004) and Mott et al. (2008) which recruited children and adolescents.

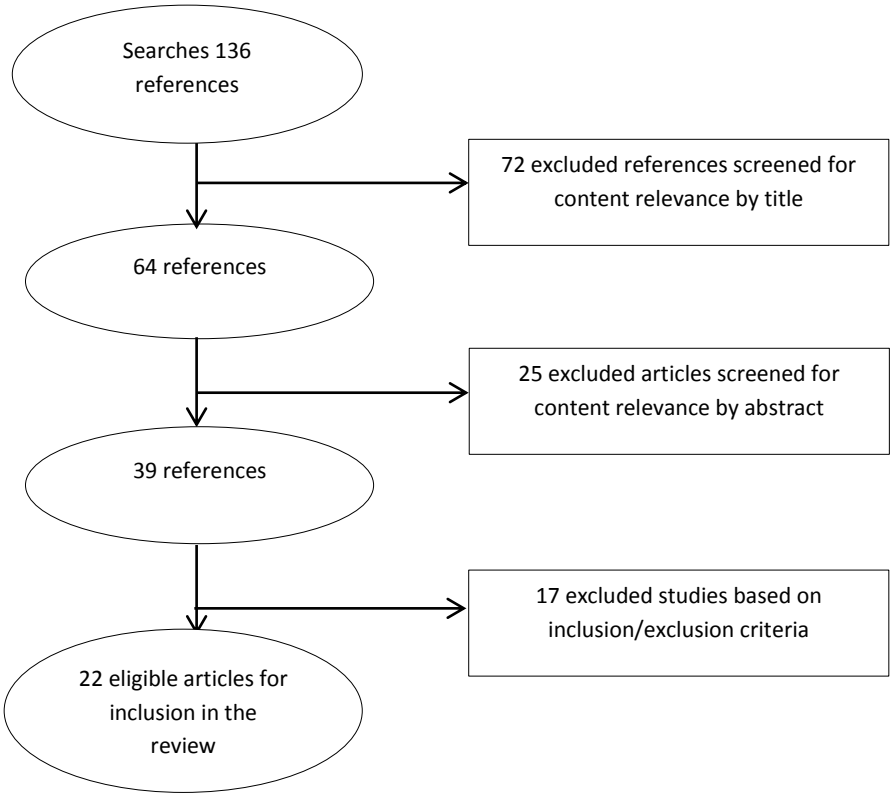


Figure 1: Flow diagram of electronic search strategy

Table 3: Randomised controlled trials of virtual reality and comparative treatment efficacy

Author (year)	Condition/problem	Sample/therapist characteristics	VR treatment characteristics	Comparative treatment characteristics	Measurement[1]	Post-treatment outcome	Follow-up	Quality rating
ANXIETY DISORDERS								
Banos et al. (2011)	PTSD, Pathological grief (PG) & Adjustment disorders (AD)	***Volunteers*** Total *N* = 39 Female = 69.2% Age: *M* = 30.9 ***Therapists***: Doctors and PhD students experienced in CBT delivery	VR-based emotional therapy (education/ VR exposure/cognitive restructuring/RP) **Duration**: 4-9 x 60-90 min. *n* = 19	Standard CBT (education/ imagined exposure/ in vivo exposure/ restructuring/RP) **Duration**: 4-9 x 60-90 min *n* = 20	***Diagnosis****:* Clinical interview using DSM-IV criteria ***Symptomatology***: FAS, BDI, PANAS, MS, Interference/Severity Scale: adapted from ADIS-IV	Both treatment groups showed significant improvements on all outcome measures. VR-therapy had significantly better outcomes than CBT on several measures (e.g. BDI).	None	21
Botella et al. (2007)	Panic Disorder & Agoraphobia (PDA)	***Clinic attendees*** Total *N* = 37 Female = 70.3% Age: *M* = 34.7 ***Therapists:*** Psychologists trained in CBT for PDA	VR exposure (VRE to interoceptive and external feared stimuli) with non VR education/restructuring/ RP **Duration**: 9 x 60 min *n* = 12	**(a)** In-vivo exposure (IVE to interoceptive and external feared stimuli) with education/RP **Duration**: 9 x 60 min *n* = 12 **(b)** waiting-list (WL) *n* = 13	***Diagnosis***: ADIS-IV ***Symptomatology***: FAS, PDSS, ASI, idiographic PA record, FQ, BDI; MS ***Treatment expectation/ satisfaction*****:** A newly designed scale	VRE and IVE groups showed significant reductions on symptomatology outcomes compared to WL. No significant difference in treatment efficacy or expectations/satisfaction between VRE and IVE.	**12 months**: Both groups continued improving and/or maintained treatment benefits. No significant differences in symptomatology or satisfaction between VR and IVE groups.	27
Choi et al. (2005)	PDA	***Clinic attendees*** Total *N* = 40 Male = 50% Age: *M* = 36.2 ***Therapists***: Unspecified characteristics	VR-CBT alongside non-VR group CBT (both components sequentially delivered education/ restructuring/ relaxation/interoceptive exposure/VR exposure) **Duration**:4 x 150 min (120 min of group CBT, 30 min VR-CBT), *n* = 20	Panic Control Program group (education/ relaxation/restructuring/ interoceptive exposure/ in vivo exposure) **Duration**: 12 x 120 min *n* = 20	***Diagnosis***: Clinical interview using DSM-IV criteria ***Symptomatology***: BDI, STAI, ASI, ACQ, BSQ, PBQ, amount of medication discontinuation	Both groups showed significant symptom reduction. No between-group differences on standardized measures or medication discontinuation.	**6 months**: Both groups maintained treatment benefits. Significantly more PCP patients had discontinued medication than VR-CBT patients; VR less efficacious that PCP long-term.	16

Emmelkamp et al. (2002)	Acrophobia	***Volunteers*** Total *N* = 33 Male = 54.5 Age: *M* = 43.97 ***Therapists***: Clinical psychology students/clinical psychologist with advanced training in behavioural therapy	VRE (to 3 feared scenarios) **Duration**: 3 x 60 min *n* = 17	IVE (to 3 scenarios) **Duration**: 3 x 60 min *n* = 16	***Diagnosis***: SCID-I, SCL-90 ***Symptomatology***: AQ, ATHQ, BAT	Both VRE and IVE significantly reduced anxiety and avoidance on all measures. No significant differences in treatment efficacy between VRE and IVE.	**6 months**: Symptom reduction maintained for both groups. No significant difference in treatment efficacy between VRE and IVE groups.	16
Krijn et al. (2007)	Aviophobia (FOF)	***Volunteers*** Total *N* = 59 Female = 66% Age *M* = 38.58 ***Therapists***: Unspecified characteristics	VRE followed by non-VR group-based CT (GBCT: education/cognitive techniques) **Duration**: 9 sessions (4 x 60 min VRE, 5 x 60 min GBCT) *n* = 30	**(a)** CBT (education/ relaxation/IVE) followed by GBCT (education/ cognitive techniques) **Duration**: 4-9 sessions (2-4 x 60 min CBT, 5 x 60 min GBCT) *n* = 23 **(b)** Bibliotherapy (BIB: educational booklet) **Duration**: 5 x 60 min *n* = 19	***Diagnosis***: The MINI, SCL-90 ***Symptomatology***: FAS, FAM *Coping*: CERQ-F, and a newly designed self-efficacy questionnaire.	VRE and CBT groups showed reduced scores on the FAS and FAM compared to BIB. VRE and CBT not statistically different in symptom reduction. CBT group showed more positive coping cognitions than VRE group. The addition of GBCT had less effect on VRE than CBT efficacy. CBT followed by GBCT was most efficacious treatment.	None	17
McLay et al. (2011)	PTSD	***Clinic attendees*** (active military servicemen) Total *N* = 20 Male: 95% Age *M* = 28.4 ***Therapists*:** Licensed Psychologists	VR gradual exposure therapy (VRGET) with non-VR education/ attentional control/ relaxation training/hotspot identification **Duration**: 10-20 x 60 min *n* = 10	TAU (including: CBT, EMDR, prolonged exposure and medication) **Duration**: varied *n* = 10	***Diagnosis***: The MINI, The CAPS ***Symptomatology***: The CAPS	VRGET group showed significant reduction in symptomatology compared to TAU.	None	18

Michaliszyn et al. (2010)	Arachnophobia	***Students*** Total *N* = 43 Female = 98% Age *M* = 29 ***Therapists***: Doctoral students	VRE (to spiders) with non-VR education/cognitive restructuring/RP **Duration**: 8 x 90 min *n* = 16	**(a)** In vivo exposure (IVE) with education/ cognitive restructuring/RP **Duration**: 8 x 90 min *n* = 16 **(b)** WL, *n* = 11	***Diagnosis***: SCID-I ***Symptomatology***: FSQ, SBQ, BAT of tarantula	Statistically and clinically significant improvements observed in both VRE and IVE compared to WL. IVE group showed significantly greater reduction on SBQ than VRE indicating reduction in problematic beliefs about spiders.	**3 months**: Benefits maintained in both groups compared to WL.	21
Mühlberger et al. (2003)	Aviophobia	***Volunteers*** Total *N* = 47 Female: 74% Age *M* = 42.2 ***Therapist***: Psychologist with advanced training in CBT	**(a)** VRE with motion simulation (MS) with non-VR CT (educational booklet/CT techniques) **Duration**: CT 1 x 60, VRE 4 x 18 min *n* = 13 **(b)** VRE without MS following CT **Duration**: 5 sessions (1 x 60 min CBT, 4 x 18 min VRE) *n* = 13	**(a)** CT (educational booklet/CT techniques) **Duration**: 1 x 60 min *n* = 11 **(b)** WL, *n* = 11	***Diagnosis***: Clinical interview using DSM-IV criteria ***Symptomatology*** (a) Primary: FFS, GFFQ, BAT of flight reservation and taking a flight (b) Secondary: FGSQ, ASI, STAI	Only VRE conditions were associated with symptom improvement; significantly better than CT and WL. Motion simulation did not enhance effectiveness of VRE. The most effective ingredient in VRE was exposure to visual and acoustic stimuli.	**6 months**: 62% of VRE groups completed actual flights, compared to 50% WL and 45% CT group; differences NS.	17
Perez-Ara et al. (2010)	PDA	***Clinic attendees*** Total *N* = 29 Female: 79.3% Age *M* = 32.8 ***Therapists***: characteristics unspecified	VR interoceptive exposure (VRIE) with non-VR CBT (education/cognitive restructuring/VR-based exposure to feared situations) **Duration**: maximum 8 x 60 min *n* = 14	Traditional interoceptive exposure (TIE) with CBT (education/cognitive restructuring/VR-based exposure to feared situations) **Duration**: maximum 8 x 60 min *n* = 15	***Diagnosis***: ADIS-IV ***Symptomatology***: FAS, ASI, PDSS	Both treatments equally effective in reducing PDA symptomatology on all measures. No significant between-group effects.	**6 months:** Treatment gains maintained or continued improving.	16
Robillard et al. (2010)	Social anxiety disorder	***Clinic attendees*** Total *N* = 45 Female: 71% Age *M* = 34.9	VRE (to social situations) with non-VR CBT **Duration**: 16 x unknown length	**(a)** IVE (to social situations) with CBT **Duration**: 16: x unknown length	***Diagnosis***: Clinical interview using DSM-IV criteria ***Symptomatology***: LSAS, SPD, ASCS, FNE, BDI, STAI, Self-	Both active treatments significantly reduced symptoms on majority of outcome measures compared	None	10

		Therapists: characteristics unspecified	*n* = 14	*n* = 16 **(b)** WL *n* = 15	efficacy single item measure	to WL. No significant differences in efficacy observed between active treatments.		
Rothbaum et al. (2006)	Aviophobia	***Volunteers*** Total *N* = 75 Female: 89% Age *M* = 39.6 ***Therapists***: characteristics unspecified	VRE with non-VR anxiety management training (AMT) **Duration**: 8 sessions (VRE 4 x 60 min, AMT 4 x 60 min) *n* = 29	**(a)** IVE with AMT **Duration**: 8 sessions (IVE 4 x 60 min, AMT 4 x 60 min) *n* = 29 **(b)** WL *n* = 17	***Diagnosis***: SCID-I ***Symptomatology***: FFI, QAF, BAT to book flight and go on actual flight. ***Treatment satisfaction*****:** CSQ	Both active treatments significantly superior to WL on willingness to fly, anxiety ratings and treatment satisfaction. No significant differences between active treatments on outcome measures or satisfaction.	**6 months:** Treatment benefits equivalent and maintained in both active treatment groups. **12 months**: Treatment benefits maintained in both groups. Participants in active treatment groups were significantly more likely to have taken subsequent flights than WL.	25
Tortella-Feliu et al. (2011)	Aviophobia	***Volunteers*** Total *N* = 60 Female: 58.3% Age *M* = 37 ***Therapists***: unspecified characteristics	VRE with non-VR CT (education/cognitive restructuring) **Duration**: maximum 6 x 60 min *n* = 19	**(a)** Computer-assisted exposure (CAE) and CT (education/cognitive restructuring) *with* therapist **Duration**: maximum 6 x 60 min *n* = 21 **(b)** CAE and CT *without* therapist. **Duration**: maximum 6 x 60 min *n* = 20	***Diagnosis***: ADIS-IV ***Symptomatology***: FFQ; FFS; clinician ratings of symptom severity ***Treatment satisfaction***: CAS	All treatments yielded large effect sizes for symptomatology reduction. No significant differences between three treatment groups on symptomatology outcomes. VRE participants rated higher levels of approval and satisfaction with treatment.	**12 months**: Treatment benefits maintained in all treatment groups. No longer any differences between groups in levels of treatment approval or satisfaction.	20
Wallach et al. (2009)	Public Speaking Anxiety (PSA)	***Students*** Total *N* = 88	VRE with non-VR CT (education/ self-	**(a)** Traditional CBT (education/self-	***Diagnosis****:* Clinical interview using DSM-IV criteria	VR-CBT and traditional CBT equally effective in reducing	None	18

		Female: 77.3% Age *M* = 27 ***Therapists***: Clinical Psychology students with training in treatment protocol	monitoring/cognitive restructuring/homework) **Duration**: 12 x 60 min (7 of these included VRE) *n* = 28	monitoring/cognitive restructuring/IVE/ homework) **Duration**: 12 x 60 min *n* = 30 **(b)** WL, *n* = 30	***Symptomatology***: LSAS, SSPS, FNE, BAT presentation to strangers	PSA compared to WL. Large effect sizes for anxiety reduction were obtained in both active treatments. Clinically significant remission was observed on for both VRCBT and CBT on avoidance scales but not fear scales. Twice as many participants dropped out of the CBT group than VR-CBT.		
Wiederhold et al. (2002)	Aviophobia	***Volunteers*** Total *N* = 30 Female: 60% Age *M* = 39.8 ***Therapists***: characteristics unspecified	**(a)** Graded VRE *without* PF following non-VR education and relaxation training **Duration**: 8 sessions (2 x 45 min education/relaxation, 6 x 30 min VRE) *n* = 10 **(b)** Graded VRET *with* physiological feedback (PF) following non-VR education and relaxation training **Duration**: education/ relaxation: 2 x 45 min, VRE: 6 x 30 min *n* = 10	Imaginal Exposure Therapy (IET) following education and relaxation training **Duration**: 8 sessions (2 x 45 min education/ relaxation,6 x 30 min IET) *n* = 10	***Diagnosis:*** Clinical interview using DSM-IV criteria ***Symptomatology*** a) Physiological: skin resistance, respiration, heart-rate (b) self-report: FOF, VAS of therapy efficacy, QAF, STAI, self-survey of stress response, SUD's of anxiety (c) behavioural: telephone call at 3 month f/u	Both VRE conditions were significantly associated with improved outcomes and superior to IET. VRE with PF was superior to VRET without, suggesting that PF strengthens VRE efficacy.	None	14
PAIN								
Gershon et al. (2004)	Procedural Pain (cancer victims)	***Hospital attendees (child cancer patients)*** Total *N* = 59 Male = 51% Age: *M* = 12.7 **Therapists:** None –	VR distraction (VRD) **Duration** 1 x 5-10 min *n* = 22	**(a)** non-VR distraction **Duration**: 1 x 5-10 min *n* = 15 **(b)** Treatment as usual (TAU) without distraction *n* = 22	***Diagnosis***: Childhood cancer ***Symptomatology***: VAS for pain intensity, CHEOPS, pulse rate, behavioural observations by nurse.	VRD group showed reduced pain and anxiety (self-report and physiologically) compared to non-VR distraction and TAU during and after procedure.	None	16

		patients trained by researcher then self-administered						
Mott et al. (2008)	Procedural pain (burn victims)	***Hospital attendees (child burn victims)*** Total *N* = 42 Male = 71% Age median = 9 ***Therapists***: None – patients trained by researcher then administered with caregiver.	Augmented Reality distraction (ARD) following analgesia and sedation. **Duration** 1 x 10-30 min (according to procedure) *n* = 20	Basic CT (relaxation/ attention-distraction/ positive reinforcement) following analgesia and sedation. **Duration**:1 x 10-30 min (according to procedure) *n* = 22	***Diagnosis***:Child burn victims ***Symptomatology***: (a) Subjective: 3-4 year olds: FLACC; 4-8 year olds: FPC-R; 8-14 year olds: VAS (b) Parent pain ratings of child's pain, respiration and pulse	ARD group showed significantly lower subjective and parental pain ratings compared to CT group. NS difference between groups on physiological measures.	None	17
Leibovici et al. (2009)	Pain: chronic non-procedural (pruritus patients)	***Hospital attendees*** (patients with dermatitis and vulgaris) Total *N* = 24 Male = 50% Age M = 44.5 ***Therapists***: None-trained by researcher then self-administered	VR distraction (VRD) **Duration**: 1 x 8 -12 min *n* = 12	Non VR distraction (audio-visual version of the VRD stimuli) **Duration**: 1 x 8-12 min *n* = 12	***Diagnosis***: Patients with pruritus related to dermatitis and vulgaris) ***Symptomatology:*** VAS of itching	No significant difference in the levels of self-reported scratching between two groups. Less scratching noted by observers in the VRD group; but the actual results of these differences were not statistically analysed.	None	17
Patterson et al. (2006)	Experimental Pain (induced thermal pain)	***Students*** Total *N* = 103 Female: 61% Age *M* = 19 ***Therapists***: None-trained by researcher then self-administered.	**(a)** VR distraction (VRD) **Duration**: 1 x 30 seconds *n* = 26 **(b)** VRD following hypnosis **Duration**: 1 x 30 seconds *n* = 26	**(a)** Hypnosis **Duration**: 1 x 30 seconds *n* = 25 **(b)** NT *n* = 26	***Diagnosis***: N/A ***Symptomatology*** (a) Primary: self-reported worst pain intensity (10-cm scale) (b) Secondary: self-reported pain unpleasantness; time spent thinking about pain; amount of 'fun' experienced during procedure (10 cm	VR significantly reduced pain regardless of suggestibility level compared to hypnosis and NT. Suggestibility moderated effects of hypnosis and VR with hypnosis. For high hypnotisability participants, VR with hypnosis showed a	None	14

					scales) ***Moderator****:* Hypnotisability-SHCS	trend towards increased efficacy in pain reduction; but interaction was NS.		
BODY-IMAGE/ OBESITY								
Manzoni et al. (2008)	Emotional distress in overeating (obesity)	***Inpatients (obesity clinic)*** Total *N* = 60 Female = 100% Age *M* = 43.67 ***Therapists***: Clinical psychologists and chartered psychotherapist	VR relaxation training (VR relaxing scenario and VRE to stress) following non-VR self-monitoring **Duration**: 12 x 60 min *n* = 20	**(a)** Imagined relaxation (imagined relaxing scenario and imagined exposure to stress) with self-monitoring **Duration**: 12 x 60 min *n* = 20 **(b)** WL , *n* = 20	***Diagnosis****:* BMI >30, EOQ ***Outcomes***: BDI, WELSQ, STAI, VAS of relaxation, heart rate	VR-R and IR groups showed significantly increased self-efficacy in eating control, decreased depressive/anxiety symptomatology and reduced post-session heart-rate compared to WL. No significant differences between VR and IR treatment efficacy.	None	28
Riva et al. (2006)	Body-image distortion & obesity	***Clinic attendees (obesity clinic)*** Total *N* = 211 Female: 100% Age *M* = 36.1 ***Therapists***: Clinical psychologists and chartered psychotherapist	VR-CT (VR exposure and response prevention/ problem-solving relapse mechanisms/VR body image comparisons/ building coping strategies and emotional awareness) following non-VR Nutritional group (NG) **Duration**: 20 x 60 (VR-CT 15 x 60 min, NT 5 x 60 min) *n* = 56	**(a)** CBT group (self-monitoring/cognitive restructuring/goal setting) following NG **Duration**: 20 x 60 (CBT 15 x 60 min, NG 5 x 60 min). *n* = 52 **(b)** Nutritional group (NG; self-monitoring/ education/physical training/diet) **Duration**: 5 x 60 min *n* = 50 **(c)** WL, N = 53	***Diagnosis****: BMI>40* ***Outcomes***: Subjective: DIET, STAI, WELSQ, BIAQ, CDRS Objective: Amount of weight loss	All active treatments associated with significant weight reduction and improved psychological outcomes compared to WL. Weight loss slightly greater in CBT group compared to NT and VR-CT but difference was NS. VRCT and CBT both superior to WL on all obesity related outcomes.	**6 months**: All three treatments maintained benefits though NT group were slightly heavier at post-treatment (NS). VR-CT superior to CBT, NT and WL in greater body image satisfaction and self-efficacy. VR-CT group also showed trend towards greater weight loss compared to NT and CBT but differences were NS.	22
Riva et al. (2001)	Body-image distortion in	***Inpatients (obesity clinic)***	VR-CT (VR exposure/ response prevention/ VR	CBT-based Psycho-nutritional	***Diagnosis***: BMI > 35, EDI-2 ***Outcomes***: DIET, STAI, AI,	Both active treatments associated with improved	None	26

	obesity	Total *N* = 28 Female: 100% Age *M* = 31.2 ***Therapists***: chartered psychologists and psychotherapist	presented desired miracle and body image distortions) with non-VR low calorie diet and physical training **Duration**: 7 x 50 min *n* = 14	group (stress management/problem solving/education) with low calorie diet and physical training **Duration**: 7 x 50 min *n* = 14	WELSQ, URICA, BSS, BIAQ, FRS, CDRS	outcomes. VR-CT significantly more effective than CBT nutritional group at improving body-satisfaction, self-efficacy, motivation and reduced over eating. VR-CT participants showed trend towards greater weight loss than CBT group, but differences NS.		
Riva et al. (2002)	Body-image distortion in Binge Eating Disorder	***Clinic attendees (obesity clinic)*** Total *N* = 20 Female: 100% Age *M* = 30.3 ***Therapists***: Clinical psychologists and chartered psychotherapist	VR-CT (exposure/response prevention/viewing desired miracle/body image distortions) **Duration**: 7 x 50 min *n* = 10	CBT-based nutritional group (stress management/problem solving/education) **Duration**: 7 x 50 min *n* = 10	***Diagnosis***: DSM-IV criteria ***Outcomes***: EDI-2, DIET, STAI, AI, WELSQ, URICA, BSS, BIAQ, FRS, CDRS	VRT group showed significantly greater improvements on measures of body-satisfaction and self-efficacy compared to CBT, but no differences in amount of binge-eating behaviour between treatments. Both treatments associated with significant reductions in binge eating.	None	24

[1]*Diagnostic measures*: ADIS-IV (Anxiety Disorders interview Schedule-4th Edition), SCID-I (Structured Clinical Interview for DSM-IV Disorders), SCL-90 (Symptom Checklist-90), MINI (Mini International Neuropsychiatric Interview), CAPS (Clinical Administered PTSD Scale), EDI-2 (Eating Disorders Inventory-2nd Edition)
Symptomatology measures: FAS (Fear & Avoidance Scales), BDI (Beck Depression Inventory), PANAS (Positive and Negative Affect Schedule), MS (Maladjustment Scale), PDSS (Panic Disorder Severity Scale), ASI (Anxiety Sensitivity Index), FQ (Fear Questionnaire), STAI (State-Trait Anxiety Inventory), ACQ (Agoraphobic Cognition Questionnaire), BSQ (Body Sensations Questionnaire), PBQ (Panic Belief Questionnaire), AQ (Acrophobia Questionnaire), ATHQ (Attitudes Towards Heights Questionnaire), BAT (Behavioural Avoidance Test), FAS (Flight Anxiety Situations Questionnaire), FAM (Fight Anxiety Modality Questionnaire), CERQ-F (Cognition Emotion Regulation Questionnaire–Flying), FSQ (Fear of Spiders Questionnaire), SBQ (Spider Beliefs Questionnaire), GFFQ (General Fear of Flying Questionnaire), PDSS (Panic Disorder Severity Scale), LSAS (Liebowitz Social Anxiety Scale), SPS (Social Phobia Scale), ASCS (Appraisal of Social Concerns Scale) FNE (Fear of Negative Evaluation), FAF (Fear of Flying Inventory), QAF (Questionnaire on Attitudes Towards Flying), SSPS (Self-Statements during Public Speaking), FNE (Fear of Negative Evaluation), FOF (Fear of Flying Inventory), QAF (Questionnaire on Attitudes Towards Flying), VAS (Visual Analogue Scale) CHEOPS (Children's Hospital of Eastern Ontario Pain Scale), FLACC (The Faces, Legs, Activity, Cry and Consolability), FPC-R (Faces Pain Scale-Revised), EOQ (Emotional Overeating Questionnaire), WELSQ (Weight Efficacy Life-Style Questionnaire), DIET (The Dieters Inventory of Eating Temptations), BIAQ (Body Image Avoidance Questionnaire), CDRS (Contour Drawing Rating Scale), AI (Assertiveness Inventory), URICA (University of Rhode Island Change Assessment Scale), BSS (Body Satisfaction Scale), RFRS (Ray Figure Rating Scale)
Secondary measures: FGSQ (Fear and General Symptoms Questionnaire), CAS (Credibility/Acceptance Scales), SHCS (Stanford Hypnotic Clinical Scale)

3.2.3 Participant characteristics

Anxiety disorders: The total number of participants included from the 14 anxiety disorders studies was 645. All participants were adults, with an age range of 17 to 72 years. Five studies recruited from clinic/hospital settings (Botella et al., 2007; Choi et al., 2005; McLay et al., 2011; Perez-Ara et al., 2010; Robillard et al., 2010) and the remainder from student and community volunteer samples. Females typically outnumbered males, usually by 2:1. The majority of participants were being treated for specific phobias (n = 435), of which 271 presented with aviophobia, 88 with public speaking anxiety, 43 with arachnophobia and 33 with acrophobia. Of the remaining participants, 106 presented with panic disorder/agoraphobia, 45 with social anxiety disorder, and 59 with PTSD.

Participants were diagnosed with anxiety disorders using either a well-established diagnostic semi-structured interview, or by clinical interview/consultation of DSM-IV criteria. Specifically, three studies (Botella et al., 2007; Tortella-Feliu et al., 2011; Perez-Ara et al., 2010) used the Anxiety Disorders Interview Schedule-IV (ADIS-IV; Brown, DiNardo & Barlow, 1994), two studies (Krijn et al., 2007; McLay et al., 2011) used the Mini International Neuropsychiatric Interview version 6 (the MINI 6; Sheehan & Lecrubier, 1990) and three studies (Emmelkamp et al., 2002; Michalisyn et al., 2010; Rothbaum et al., 2006) used the Structured Clinical interview for DSM-IV Axis I Disorders (SCID-I; First, Spitzer, Gibbon & Williams, 2002). The remaining six studies (Banos et al., 2011; Choi et al., 2005; Muhlberger et al., 2003; Robillard et al., 2010; Wallach et al., Wiederhold et al., 2002) diagnosed participants using clinical interviews based on DSM-IV criteria.

With the exception of Robillard et al. (2010), all anxiety disorders studies reported exclusion criteria, and there was considerable overlap in these criteria across studies (e.g. psychotic illness, drug/alcohol abuse, epilepsy). Of those

participants assessed for eligibility, 5% (Rothbaum et al., 2006), 24% (Banos et al., 2011), 20% (Botella et al., 2007), 35% (Emmelkamp et al., 2007), 80% (Muhlberger et al., 2003) and 30% (Tortella-Feliu et al., 2011) were excluded. The numbers of excluded participants were not reported in the remaining anxiety disorders studies.

Pain: The total number of participants included from the pain studies was 228. Of these, 127 were adults and 101 were children/adolescents. The gender split across trials was roughly equal. Participants' ages ranged from 3.5 (Mott et al., 2008) to 84 years (Leibovici et al., 2009). One-hundred-and-three participants were being treated for experimental pain, 101 for procedural pain and 24 for chronic pain. The majority of participants were recruited from hospitals (n = 104) and the remainder were students (n = 24). Fifty-nine participants had a diagnosis of cancer (Gershon et al., 2004), 42 had burn injuries (Mott et al., 2008), and 24 presented with chronic skin conditions (Leibovici et al., 2009). Participants undergoing experimental pain (Patterson et al., 2006) were student volunteers with no diagnoses.

Two pain studies reported exclusion criteria (Gershon et al., 2004; Leibovici et al., 2009) and two did not (Mott et al., 2008; Patterson et al., 2006). Only Leibovici et al. (2009) reported the percentage of participants that were excluded (11%) from taking part.

Body-image distortion/obesity: The total number of participants included from body-image distortion/obesity trials was 319. All were adult female inpatients and outpatients of obesity clinics. Age ranges were not reported. The mean ages of participants were 36.1 (Riva et al., 2006), 30.5 (Riva et al., 2002), 30.3 (Riva et al., 2001) and 43.7 (Manzoni et al., 2008). All participants were being treated for psychological variables related to episodes of overeating, `including body-image distortion (n = 259) and emotional stress/distress (n = 60).

Two-hundred-and-ninety-nine participants were classified as obese according to body mass index and 20 were diagnosed with binge eating disorder according to DSM-IV criteria. Three of the four body-image distortion/obesity studies reported exclusion criteria (Riva et al., 2001; Riva et al., 2002; Riva et al., 2006). Only Riva et al. (2006) reported the percentage of participants that were excluded from taking part (28%).

3.2.4 VRT intervention characteristics

Considering all studies in combination, VRTs consisted of between one (Gershon et al., 2004; Leibovici et al., 2009; Mott et al., 2008; Patterson et al., 2006) and 20 sessions (McLay et al., 2011). Session length varied from 30 seconds (Patterson et al., 2006) to 150 minutes (Choi et al., 2005). These treatment times represent an aggregate of all VRT components (e.g. both VR and therapist-delivered components).

Anxiety disorders: The anxiety disorders VRTs had between three (Emmelkamp et al., 2003) and 20 sessions (McLay et al., 2011), varying from 30 minutes (Wiederhold et al., 2002) to 150 minutes (Choi et al., 2005). Robillard et al. (2010) did not report session duration. With the exception of Emmelkamp et al. (2002) who delivered VRET in isolation, all VRT interventions included therapist-facilitated cognitive therapy (e.g. education, cognitive restructuring) alongside VRET. In most cases, cognitive therapy was delivered during VRET sessions, but in four studies it was conducted in separate sessions (Choi et al., 2005; Krijn et al., 2007; Rothbaum et al., 2006; Wiederhold et al., 2002).

The majority of anxiety disorder studies (*n* = 12) included one VRT intervention condition, whilst two studies assigned participants to one of two VRT interventions. Muhlberger et al. (2003) had two VRET interventions for aviophobia;

one with motion simulation and one without. Wiederhold et al's. (2002) two VRT interventions for aviophobia consisted of VRET with, and without physiological feedback.

In terms of the therapist characteristics, two studies used a combination of doctoral level students and qualified doctors/psychologists (Banos et al., 2011; Emmelkamp et al., 2002), three used clinical psychologists/psychotherapists trained in CBT, or CBT-therapists (Botella et al., 2007; McLay et al., 2011; Muhlberger et al., 2003) and two used doctoral level students (Michaliszyn et al., 2010; Wallach et al., 2009). Seven studies did not report therapists' characteristics (Choi et al., 2005; Krijn et al., 2007; Perez-Ara., 2010; Robillard et al., 2010; Rothbaum et al., 2006; Tortella-Feliu et al., 2011; Wiederhold et al., 2002).

Pain: VRTs for pain management were single sessions, ranging in length from 30 seconds (Patterson et al., 2006) to 30 minutes (Mott et al., 2008). All interventions consisted of VR distraction and were self-administered following brief training on equipment use by a researcher. Gershon et al. (2004), Mott et al. (2008) and Leibovici et al. (2009) included one VRT condition. Patterson et al. (2006) assigned participants to either VR distraction in isolation, or VR distraction following hypnosis.

Body-image distortion/obesity: VRTs for body-image distortion/obesity had between seven (Riva et al., 2001; Riva et al., 2002) and 20 sessions (Riva et al., 2006), ranging in length from 50 minutes (Riva et al., 2001; Riva et al., 2002) to 60 minutes (Manzoni et al., 2008; Riva et al., 2006). All four interventions were based on CBT treatment principles and included exposure to overeating triggers and response prevention training. Manzoni et al. (2008) also included VR-based relaxation training and therapist-assisted self-monitoring, conducted in separate sessions. The remaining three studies (Riva et al., 2001; Riva et al., 2002; Riva et al., 2006) included VR-based body image comparisons/distortions and accompanying

cognitive techniques, facilitated by therapists during VRT. Therapists across studies were either clinical psychologists or chartered psychotherapists.

3.2.5 Comparative intervention characteristics

Anxiety disorders: Comparative treatments had between one (Muhlberger et al., 2003) and sixteen (Robillard et al., 2010) sessions. Session length varied from 30 minutes (Wiederhold et al., 2002) to 120 minutes (Choi et al., 2005). Most comparative interventions (*n* = 11) were equal in number and length of sessions to their VRT counterparts. This was with the exception of Choi et al. (2005) in which the length of comparative treatment was greater than that of VRT, and Muhlberger et al. (2003) for which the opposite was true. McLay et al. (2011) did not report the number and length of comparative treatment sessions.

All comparative anxiety disorder interventions were based on cognitive and/or behavioural treatment. With the exception of Muhlberger et al. (2003), all included in vivo or interoceptive exposure (in the case of panic disorder). Muhlberger et al.'s (2003) intervention consisted of purely cognitive techniques and psychoeducation. Conversely, Emmelkamp et al.'s (2002) comparative intervention was exclusively exposure-based. Comparative interventions often included psychoeducation, cognitive restructuring, relaxation training and relapse prevention strategies.

Six studies had one comparative intervention (Banos et al., 2011; Choi et al., 2005; Emmelkamp et al., 2002; McLay et al., 2011; Perez-Ara et al., 2010; Weiderhold et al., 2002). Two studies included two comparative treatments: Krijn et al. (2007) assigned participants to either CBT or bibliotherapy; Tortella-Feliu et al. (2011) delivered computer-assisted exposure and cognitive therapy either in the presence of the therapist, or via self-help. McLay et al.'s comparative treatment condition was treatment-as-usual (TAU) but the number of participants who received

different components of TAU (CBT, EMDR, exposure and medication) was not reported. Six studies included a no-treatment/wait-list condition in addition to a comparative treatment (Botella et al., 2007; Michaliszyn et al., 2010; Muhlberger et al., 2003; Robillard et al., 2010; Rothbaum et al., 2006; Wallach et al., 2009). Therapists' characteristics across the studies were synonymous with those involved in VRT treatment delivery.

Pain: The comparative interventions were single sessions, varying in length between 30 seconds (Patterson et al., 2006) and 30 minutes (Mott et al., 2008). Gershon et al. (2004) and Leibovici et al. (2009) used distraction-based treatments. Mott et al. (2008) also included distraction in their broader CBT comparative intervention. Patterson et al.'s (2006) intervention was hypnosis. All comparative interventions were equal in number and length to their VRT counterparts and were self-administered. Gershon et al. (2004) and Patterson et al. (2006) included additional no-treatment conditions.

Body-image distortion/obesity: Comparative treatments were between seven (Riva et al., 2001; Riva et al., 2002) and 20 sessions (Riva et al., 2006), ranging in length between 50 minutes (Riva et al., 2001; Riva et al., 2002) and 60 minutes (Riva et al., 2006; Manzoni et al., 2008). Three studies (Riva et al., 2001; Riva et al., 2002; Riva et al., 2006) delivered group CBT and one study (Manzoni et al., 2008) taught relaxation and self-monitoring practices. Riva et al. (2001) and Riva et al. (2002) included one comparative intervention, whilst Riva et al. (2006) assigned participants to either a CBT or nutritional group. Two of the four studies also included no-treatment conditions (Manzoni et al., 2008; Riva et al., 2006). Therapists' characteristics were the same as those involved in VRT delivery.

3.2.6 Outcome measures characteristics

Anxiety disorders: All anxiety disorders studies used well validated self-report measures of primary symptomatology, with the majority administering four or more different measures. Two studies included validated secondary outcome measures (Muhlberger et al., 2003; Patterson et al., 2006), whilst others indexed behavioural (e.g. a behavioural avoidance test) and physiological outcomes. Three studies also administered self-report measures of treatment satisfaction (Botella et al., 2007; Rothbaum et al., 2006; Tortella-Feliu et al., 2011).

Eight anxiety disorder trials measured long-term outcomes; Michaliszyn et al. (2010) took measurements at three-months. Four studies assessed outcomes at six-months (Choi et al., 2005; Emmelkamp et al., 2002; Muhlberger et al., 2003; Perez-Ara et al., 2010), two at 12 months (Botella et al., 2007; Tortella-Feliu et al., 2011) and one study obtained outcomes at both six and 12 months (Rothbaum et al., 2006).

Pain: All four studies used scales (Likert or visual analogue) of pain intensity as outcome measures. Three of these were subjective scales completed by the participants (Gershon et al., 2004; Leibovici et al., 2009; Patterson et al., 2006) and one obtained parental estimates (Mott et al., 2008). Two studies also administered validated self-report primary outcome measures (Gershon et al., 2004; Mott et al., 2008). Long-term follow-up outcomes were not obtained in any of the pain studies.

Body-image distortion/obesity: All studies used multiple outcome measures of symptomatology. Three studies (Riva et al., 2001; Riva et al., 2002; Riva et al., 2006) included measures of body image satisfaction and distortion, and all four administered the weight efficacy lifestyle questionnaire (Clark, Abrams, Niaura, Etton & Rossi, 1991). Riva et al. (2006) also included weight loss as a primary

outcome. Additionally, all studies administered validated self-report measures of mood (e.g. the State-Trait Anxiety Inventory; Spielberger, 1968). Manzoni et al. (2008) used physiological measurement in addition to self-report tools. Riva et al. (2006) measured outcomes at six-month follow-up.

3.3 Risk of bias of included studies

3.3.1 Sequence generation

Anxiety disorders: Two studies adequately described the method of allocation, with no bias possible (Wiederhold et al., 2002; McLay et al., 2011). The remaining 12 studies did not provide full details of the method of allocation and some bias was believed to be possible from the descriptions given.

Pain: Two studies adequately reported the allocation method and were therefore judged to be free of bias (Gershon et al., 2004; Mott et al., 2008). Bias was possible in the remaining two studies (Leibovici et al., 2009; Patterson et al., 2006).

Body-image distortion/obesity: Manzoni et al. (2008) and Riva et al. (2006) reported the method of allocation such that no bias was possible. Riva et al. (2001) and Riva et al. (2002) did not report details of allocation.

3.3.2 Allocation concealment

Anxiety disorders: Most studies did not provide full details of the method of randomisation and therefore concealment was unclear or inadequate in 12 studies. There was reporting of adequate concealment procedures in two studies (McLay et al., 2011; Wiederhod et al., 2002).

Pain: The method of concealment was not described in any of the four pain studies and therefore bias was possible.

Body-image distortion/obesity: Only Riva et al. (2006) described the method of allocation concealment sufficiently.

3.3.3 Blinding

A double blind methodology for studies of psychological treatment is impossible as it is clear to participants what treatment they are receiving. However, a well-designed study should ensure blinding of the assessor of outcome measures.

Anxiety disorders: Blinding of assessors was clearly demonstrated in five of the anxiety disorder studies (Banos et al., 2011; Botella et al., 2007; McLay et al., 2011; Rothbaum et al., 2006; Tortella-Feliu et al., 2011).

Pain: Of the four pain studies, only Patterson et al. (2006) attempted to blind assessors to treatment conditions.

Body-image distortion/obesity: All four studies (Manzoni et al., 2008; Riva et al., 2001; Riva et al., 2002; Riva et al., 2006) clearly demonstrated blinding of assessors.

3.3.4 Loss to follow-up

Anxiety disorders: Attrition rates were reported in all except one study (Robillard et al., 2010). There was no attrition in three studies (McLay et al., 2011; Perez-Ara et al., 2010; Wiederhold et al., 2002). The highest reported drop-out rate was 33% (Tortella-Feliu et al., 2011). Reasons for attrition were only sometimes reported. Krijn et al. (2007), Michaliszyn et al. (2010), Rothbaum et al. (2006) and Tortella-Feliu et al. (2011) provided explanations for drop-out and conducted intention-to-treat (ITT) analysis using last-observation carried forward. Emmelkamp et al. (2002), Muhlberger et al. (2003) and Wallach et al. (2009) described reasons for attrition but did not undertake ITT analysis. In other studies, the numbers of withdrawals were recorded but reasons were not provided.

Pain: There was no attrition in any of the pain studies.

Body-image distortion/obesity: There was no attrition in any of the studies.

3.3.5 Additional methodological and reporting issues

The quality of studies in relation to other methodological and reporting criteria was variable (see Table 4). The average level of Downs and Black (1998) criteria fulfilled by studies was 59.8%. Overall, levels of reporting were generally high (*M* = 73.1). With the exception of Mott et al. (2008) all studies were considered to have clear aims, and with the exception of Robillard et al. (2010), all had well-defined inclusion criteria and description of measures. However, estimates of random variability (e.g. standard deviations) were not provided in two pain studies (Gershon et al., 2004; Patterson et al., 2006) or in two body-image distortion/obesity trials (Riva et al., 2001; Riva et al., 2002). Also of concern is that of the 22 studies included in the review, only one (Michaliszyn et al., 2010) reported on potential adverse consequences of the intervention.

Table 4: Percentage of Downs & Black criteria fulfilled by included studies

	Quality criteria					
Comparative VRT disorder type	Reporting %	External validity %	Internal validity: bias %	Internal validity: confounding %	Power %	Overall %
Anxiety (*n* = 14)	74.0	11.9	72.4	66.7	14.3	55.4
Body-image (*n* = 4)	79.5	75.0	85.7	79.1	75.0	75.0
Pain (*n* = 4)	63.6	25.0	60.7	79.1	0.00	48.5
M (*N* = 22)	73.1	26.1	72.0	70.3	26.1	59.8

Levels of internal validity-bias and confounding were generally good (*M* = 72.0 and *M* = 70.3, respectively), indicating high scores on items such as appropriate statistical analyses, and synonymous methods of recruitment for active treatment groups. Conversely, external validity scores were low (*M* = 26.1), which was unsurprising given the inherent characteristics of RCTs. However, external validity was relatively high across body-image distortion/obesity studies, which reflects the fact that participants were mainly unselected consecutive samples, and treatment took place at the hospitals that participants normally attended.

Of note were the overall low levels of power (*M* = 26.1), which reflected the fact that this item on the Downs and Black (1998) checklist required power calculations to be performed. Only five studies demonstrated this; two were anxiety disorders (Botella et al., 2007; Rothbaum et al., 2006) and three were body-image distortion/obesity trials (Manzoni et al., 2008; Riva et al., 2001; Riva et al., 2002).

3.4 Effects of interventions

Eighteen of the 22 studies provided data available for analysis. For the four studies in which insufficient data was provided (Muhlberger et al., 2003; Gershon et al., 2004; Riva et al., 2001; Riva et al., 2002), attempts were made to contact the authors in order to obtain data but these were unsuccessful.

3.4.1 Anxiety disorders

Comparison one: Post-treatment comparative efficacy for anxiety disorders outcomes

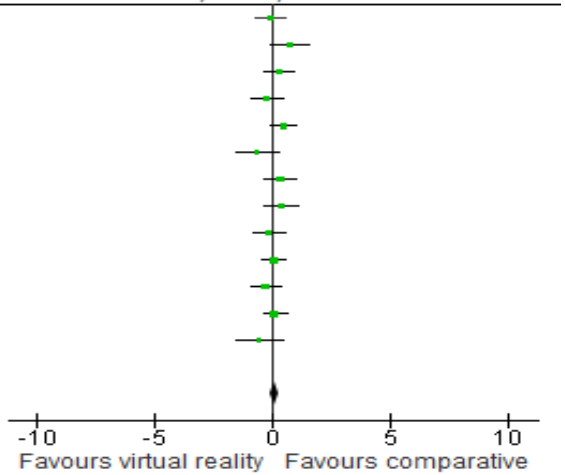
Thirteen of the 14 anxiety disorder studies were included in the analysis of post-treatment outcomes. Muhlberger et al. (2003) did not provide estimates of population variance for analysis. The primary outcome measures used in the

analysis are listed in Figure 2. Lower scores on all measures represent improvements in symptomatology.

There was no evidence of statistical heterogeneity (χ^2 = 13.03; df = 12; p = .37; I^2 = 8%), therefore a fixed effects model was used (see Figure 2). Post-treatment, there was no significant difference between VRT and comparative psychological treatment for primary symptomatology outcomes (k=13, n=473; SMD = 0.08, 95% CI -0.10-0.26, p =.40).

Study or Subgroup	Virtual reality treatment Mean	SD	Total	Comparative treatment Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Banos et al. (2011) (1)	6.5	4.57	19	7	6.4	20	8.4%	-0.09 [-0.72, 0.54]
Botella et al. (2007) (2)	14.75	5.86	12	10.67	4.54	12	4.8%	0.75 [-0.08, 1.58]
Choi et al. (2005) (3)	19.16	11.52	20	15.65	11.38	20	8.5%	0.30 [-0.32, 0.92]
Emmelkamp et al. (2002) (4)	31.18	14	17	34.25	10.66	16	7.1%	-0.24 [-0.93, 0.45]
Krijn et al. (2007) (5)	21.87	15.5	30	14.96	11.57	23	10.9%	0.49 [-0.06, 1.04]
McLay et al. (2011) (6)	48.1	36.9	10	72.3	33.8	10	4.0%	-0.66 [-1.56, 0.25]
Michaliszyn et al. (2010) (7)	54.37	22.46	16	47.88	14.07	16	6.8%	0.34 [-0.36, 1.04]
Perez-Ara et al. (2010) (8)	15.46	8.43	14	12.71	6.95	15	6.1%	0.35 [-0.39, 1.08]
Robillard et al. (2010) (9)	47.5	17.83	14	50.38	23.87	16	6.4%	-0.13 [-0.85, 0.59]
Rothbaum et al. (2006) (10)	103.69	49.35	29	100.34	43.49	29	12.5%	0.07 [-0.44, 0.59]
Tortella-Feliu (2011) (11)	109.03	40.6	19	122.4	51.5	21	8.5%	-0.28 [-0.90, 0.34]
Wallach et al. (2009) (12)	9.2	13.55	28	7.9	10.24	30	12.5%	0.11 [-0.41, 0.62]
Wiederhold et al. (2002) (13)	83.38	32.95	8	111.44	57.49	9	3.5%	-0.56 [-1.54, 0.42]
Total (95% CI)			**236**			**237**	**100.0%**	**0.08 [-0.10, 0.26]**

Heterogeneity: Chi² = 13.03, df = 12 (P = 0.37); I² = 8%
Test for overall effect: Z = 0.85 (P = 0.40)



(1) PTSD: the BDI
(2) PDA: The Anxiety Sensitivity Index
(3) PDA: The Anxiety Sensitivity Index
(4) Acrophobia: THe Attitudes towards Heights Questionnaire
(5) Aviophobia: The Flight Anxiety Situations Questionnaire
(6) PTSD: The Clinician Administered PTSD Scale
(7) Arachnophobia: The Fear of Spiders QUestionnaire
(8) PDA: The Anxiety Sensitivity Index
(9) SAD: The Liebowitz Social Anxiety Scale
(10) Aviophobia: The Fear of Flying Inventory
(11) Aviophobia: The Fear of Flying Questionnaire
(12) Public speaking anxiety: The Liebowitz Social Anxiety Scale
(13) Aviophobia: Fear of Flying Inventory

Figure 2: Post-treatment comparative efficacy for anxiety disorder outcomes

Figure 3 shows that the studies included in comparison one were distributed roughly symmetrically around the combined effect size, indicating a low risk of publication bias (Borenstein, 2005).

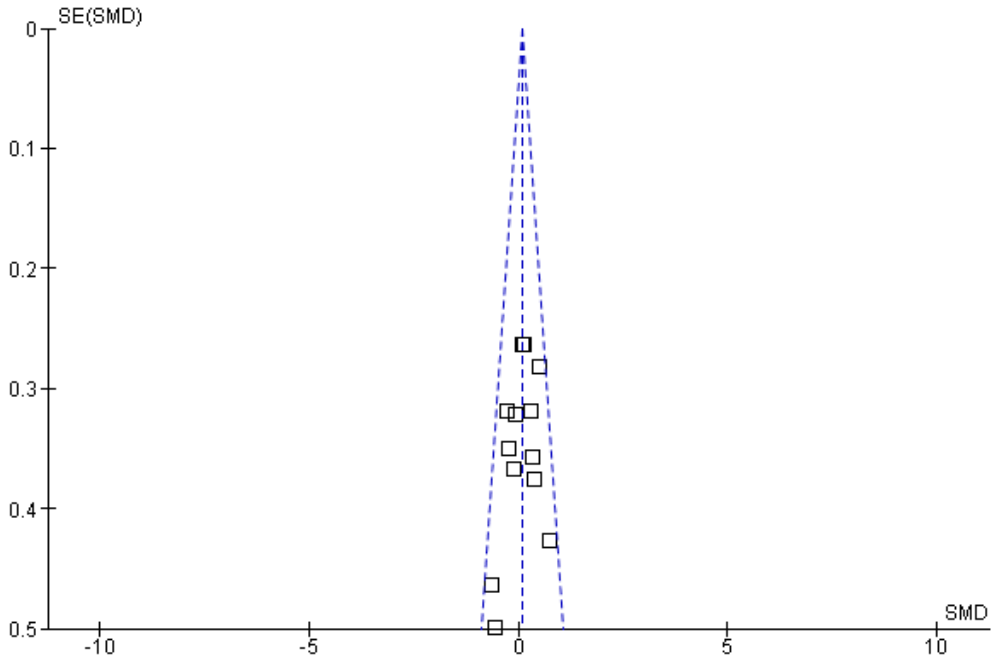


Figure 3: Funnelplot to detect for publication bias in post-treatment anxiety disorders studies

Comparison two: Long-term follow-up comparative efficacy for anxiety disorder outcomes

Six anxiety disorder studies were included in the analysis of long-term follow-up outcomes. Choi et al. (2005) did not provide sufficient numerical outcome data (means and Standard Deviations not reported) to be included in the analysis. The same primary outcome measures were used in the analysis as those selected for comparison one. Long-term follow-up measures were obtained at three months (Michaliszyn et al., 2010), six months (Emmelkamp et al., 2002; Muhlberger et al., 2003; Perez-Ara., 2010) and 12 months (Rothbaum et al., 2006; Tortella-Feliu et al., 2011). There were an insufficient number of studies to check for publication bias.

There was no evidence of statistical heterogeneity (χ^2 = 1.62; df = 5; p = .90; I^2 = 0%), therefore a fixed effects model was used (see Figure 4). At follow-up (3, 6 or 12 months), there was no significant difference between VRT and comparative psychological treatment on primary symptomatology outcome measures (k=6, n=183; SMD = 0.13, 95% CI -0.16-0.43, p = 0.37).

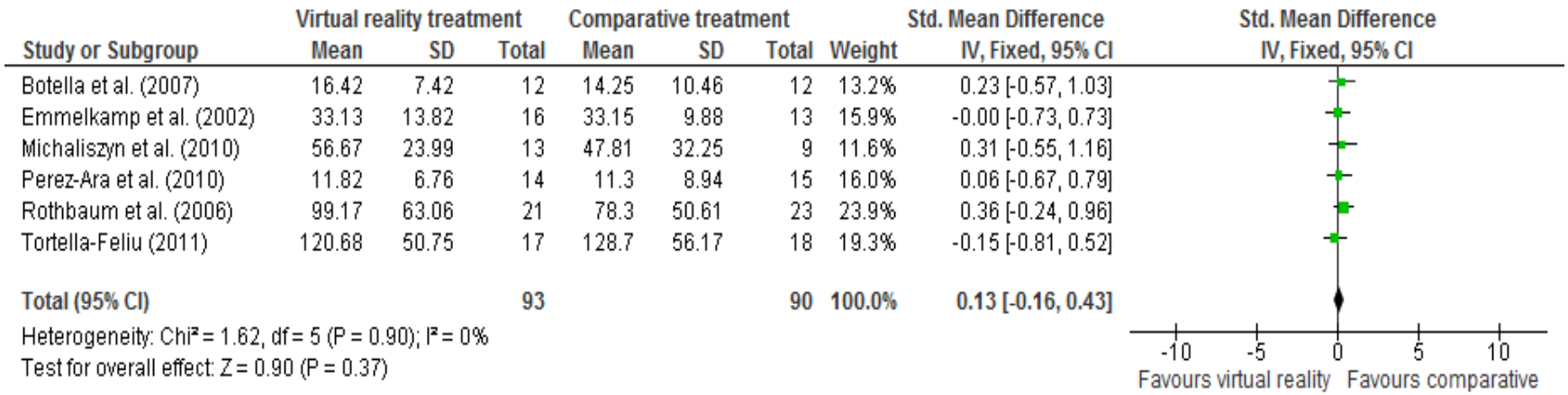
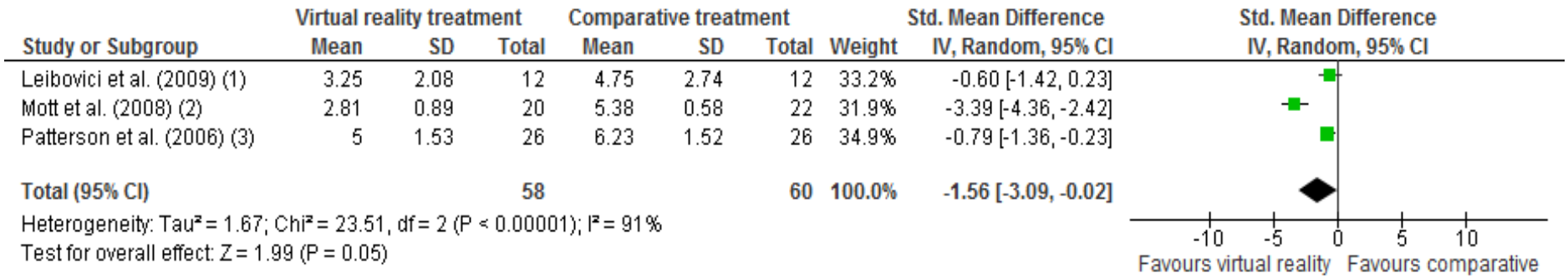


Figure 4: Long-term follow-up comparative efficacy for anxiety disorder outcomes

3.4.2. Pain

Only one comparison was run for the pain studies as follow-up data was not gathered by any of the included studies. Three of the four pain studies were included in the analysis (Liebovici et al., 2009; Mott et al., 2008; Patterson et al., 2006). Gershon et al. (2004) did not provide sufficient numerical data (no estimates of population variance) to be entered. Outcome measures selected for inclusion are detailed in Figure 5. Lower scores on all measures represent reductions in symptomatology. There were an insufficient number of studies to check for publication bias.

Statistical heterogeneity was observed ($\chi^2 = 22.51$; df = 2; $p < .01$; $I^2 = 91\%$), therefore a random effects model was used (see Figure 5). Following treatment, a significant difference was just reached between VRT and comparative psychological treatment, with VRT resulting in greater reductions of pain symptomatology (k=3, n=118; SMD = -1.56, 95% CI -3.08-0.02, $p = .05$).



(1) Chronic pain: Visual Analogue Scale for itching intensity
(2) Procedural pain: FLACC, FPS-R and Visual Analogue Scale (standardised aggregate of scores)
(3) Experimental pain: 10 cm scale for worst pain intensity

Figure 5: Post-treatment comparative efficacy for pain outcomes

3.4.3 Body-image distortion/obesity

Comparison one: Post-treatment comparative efficacy for body-image distortion/obesity outcomes

Only two of the four body-image distortion/obesity studies were included in the analysis (Manzoni et al., 2008; Riva et al., 2006). Riva et al. (2001) and Riva et al. (2002) did not provide sufficient numerical data (no estimates of population variance) for inclusion. Outcome measures selected for inclusion are detailed in Figure 6. Lower scores on all measures represent reductions in symptomatology. There were an insufficient number of trials to check for publication bias within the comparison.

There was no evidence of statistical heterogeneity ($\chi^2 = 0.06$; df = 1; $p = .80$; $I^2 = 0\%$), therefore a fixed effects model was used (see Figure 6). Following treatment, there was no difference between VRT and comparative psychological treatment on primary symptomatology outcome measures (k=2, n=150; SMD = -0.30, 95% CI -0.62-0.03, $p = .07$). There was however a trend towards increased VRT efficacy.

Comparison two: Long-term follow-up comparative efficacy for body-image distortion/obesity

Only one body-image distortion/obesity study measured outcomes at long-term (six months) follow-up (Riva et al., 2006). The inclusion of only one study precluded assessment of publication bias and heterogeneity. Since only one study was included in the comparison, checks for heterogeneity were not applicable, and a fixed effects model was used. At six-month follow-up, there was no significant difference between VRT and comparative psychological treatment on primary symptomatology outcome measures for body-image distortion/obesity (k=1, n=40; SMD = -0.62, 95% CI -1.25-0.02, $p = .06$). There was however a trend towards increased VRT efficacy, which nearly reached significance.

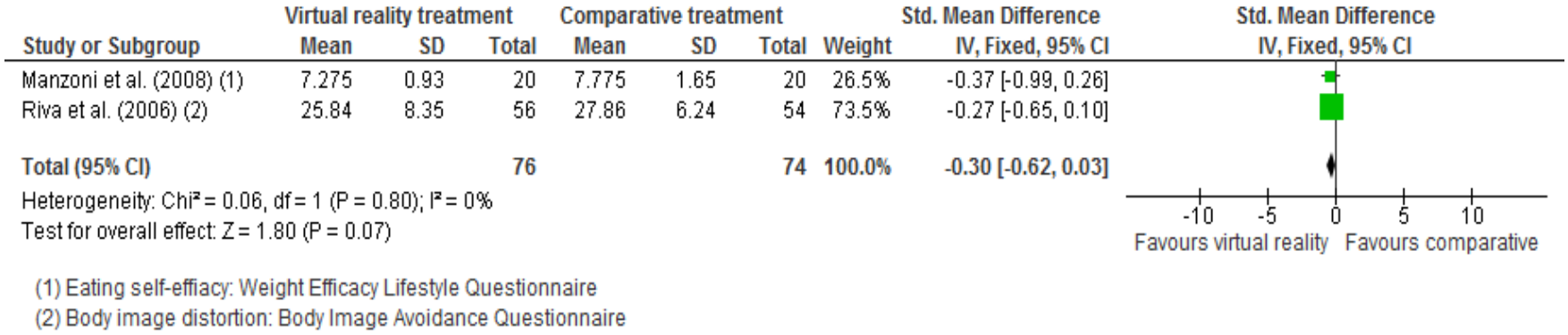


Figure 6: Post-treatment comparative efficacy for body-image distortion/obesity outcomes

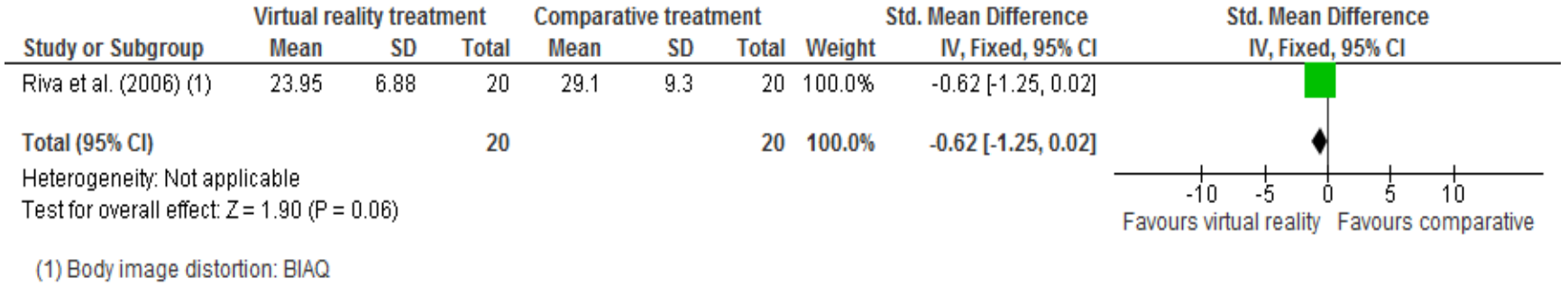


Figure 7: Long-term follow-up comparative efficacy for body-image distortion/obesity outcomes

4. DISCUSSION

4.1 Main findings

Twenty-two RCTs were identified that compared VRTs with established psychological interventions for treating anxiety disorders, body-image distortion/obesity and pain. These studies included a total of 1192 participants. Eighteen studies, with 741 participants provided data for analysis. There was no difference between VRTs and traditional treatments for anxiety disorder symptom reduction either at post-treatment or follow-up. There was an observable post-treatment difference in pain outcomes that just reached significance, favouring VRT over traditional interventions. Follow-up data for pain symptomatology was not obtained by any of the included trials. Non-significant trends were observed for body-image distortion/obesity symptomatology, favouring VRT over traditional treatments both at post-treatment and follow-up.

4.2 Overall completeness and applicability of the evidence

The studies included in this review enabled the primary research question to be addressed; it was possible to perform a meta-analysis of RCTs that compared VRTs against established therapies aimed at reducing mental health symptomatology. However, analyses could not be performed for four studies (Muhlberger et al., 2003; Gershon et al., 2004; Riva et al., 2001; Riva et al., 2002) that provided insufficient data.

The current findings tentatively indicate that VRTs are at least as effective – and in some cases superior to – traditional psychological treatments for alleviating mental health symptomatology. However, it is unclear whether traditional comparative interventions were state-of-the-art, whether they were manualised and if so, therapists' fidelity to protocols. Thus it remains possible that VRTs are comparable in efficacy to 'low grade' traditional treatments but inferior to cutting-edge counterparts. Additionally, credibility and expectancy may have influenced the

outcomes of comparisons, as these variables are known to influence treatment responsivity (Jacobson & Baucom, 1977; Kazdin, 1979; Kazdin & Krouse, 1983).

For anxiety disorder and body-image distortion/obesity symptomatology, VRTs appeared to be comparable – but not superior – to traditional interventions, both in the short- and longer-term. However, within these domains clinical conditions were heterogeneous, meaning that there remains uncertainty regarding the comparative efficacy of VRTs for specific anxiety disorders and specific obesity-related psychological outcomes. Additionally most VRTs were multifaceted and included non-VR therapeutic components. It is therefore unclear whether VRTs are efficacious as standalone treatments, or when used in conjunction with traditional treatments. Furthermore, the analyses of anxiety disorders and body-image distortion/obesity outcomes did not include the results of behavioural avoidance tests, meaning that generalisability to real life cannot be readily inferred.

The results of the pain analyses provide preliminary support for VRT as a more efficacious method of pain reduction than established psychological approaches. In addition to credibility and expectancy, there are a number of potential explanations for this observed effect. First it is possible that there was a weak response to traditional treatments due to poor design and delivery of comparative interventions. Second, VRTs may have included hidden treatment components (e.g. relaxation) over and above distraction, rendering traditional distraction an ill-matched comparison. Third, only one pain study (Patterson et al., 2006) attempted to blind outcome assessors, meaning that researcher allegiance could have influenced outcomes. Fourth, the study with the most promising data in support of VRT (Mott et al., 2008) used proxy parental pain ratings, which could be less accurate than patients' self-reports. Finally, VRT may simply be a more efficient method for consuming attentional resources than traditional distraction. These competing explanations require continued research (see implications below).

Across all three mental health domains, only one outcome measure from each study was selected for analysis; this is a limitation given that most studies administered multiple measures of symptomatology. To address this, effect sizes for all outcomes could have been aggregated using Hedges G, but this method introduces problems of its own. Only one study reported on potential adverse effects of interventions (Michaliszyn et al., 2010) and it is unclear whether adverse consequences resulted from interventions in the other studies. This is concerning in light of knowledge that VR can induce simulation-sickness and visual disturbance (Griffin, 1990). These limitations coupled with the small number of studies in most of the comparisons, their small sample sizes, and heterogeneity (see below) complicate interpretation.

4.3 Quality of evidence

Thirteen RCTs with 473 participants were included in the meta-analysis of anxiety disorders at post-treatment, suggesting that the results may be robust. However, this must be considered in light of heterogeneity and methodological quality issues. Other comparisons included fewer RCTs and participants and therefore the results are likely to be less robust, with an increased risk of chance findings. Consequently, significant findings and trends in the realms of pain and body-image distortion, respectively, should be interpreted with caution.

4.3.1 Heterogeneity

Clinical heterogeneity was noted within all three mental health domains. Although all trials attempted to reduce symptomatology, the nature of the interventions and clinical characteristics of participants were diverse. Whilst all the anxiety disorders VRT interventions included exposure, they also had varying degrees of cognitive therapy and therapist involvement. Similarly, although all traditional comparative interventions were based on CBT principles, they were

diverse in terms of treatment components, method of delivery and number/duration of sessions. It is very difficult to compare such trials and there did appear to be some substantial differences in outcomes. For example, Wiederhold et al. (2002) obtained a moderate effect size in favour of VRT, whereas Botella et al. (2007) obtained a moderate effect size in the opposite direction.

Within the analysis of pain studies statistical heterogeneity was apparent, with the I^2 value indicating inconsistencies between trials that were grouped together. Consequently a random effects model was applied in order to calculate more conservative confidence intervals. Clinical heterogeneity was also present in the pain studies, including child and adult participants with acute procedural pain, experimental pain and chronic pain. The appropriateness of aggregating these results is debatable and unfortunately it was not possible to perform sensitivity analysis owing to the small number of included trials. Nevertheless, it is encouraging that the effect sizes from all three trials analysed were medium or large, suggesting that the overall significant finding was not artificially elevated by the results of one study.

Clinical heterogeneity was also present in the two body-image distortion/obesity trials. Although both studies included women diagnosed as clinically obese, they were addressing different psychological constructs. Riva et al. (2006) specifically targeted body image distortion, whilst Manzoni et al. (2008) targeted broader eating-related self-efficacy. Furthermore the VRT and comparative interventions varied considerably between trials; Manzoni et al. (2008) relied almost exclusively on relaxation methods whereas Riva et al. (2006) delivered comprehensive CBT packages. Such heterogeneity coupled with the small combined samples sizes both at post-treatment (n = 150) and follow-up (n = 40) limit the utility of these findings.

In spite of heterogeneity across trials, all studies were attempting to measure mental health symptom reduction in response to VRT compared to established

treatments. Consequently it remains worthwhile summarising their combined results, but means that caution should be exercised when interpreting findings.

4.3.2 Methodological quality

The overall quality of the studies was variable. Several issues were apparent in many of the studies including lack of information on the randomisation process, incomplete reporting and analyses of attrition, lack of follow-up data and insufficient statistical power. The small sample sizes of many of the studies are also a notable limitation. However, the VRT and comparative intervention participants in most studies appeared well matched at baseline, reducing the risk of the reported unadjusted means being influenced by pre-treatment differences. Additionally, the overall ratings obtained on the Downs and Black (1998) checklist suggested that the methodological quality of trials was generally acceptable.

4.4 Potential biases in the review process

This review was guided by Cochrane Collaboration guidelines, which would have reduced potential bias. Methodological quality of studies was assessed using a validated tool (the Downs & Black checklist, 1998), further reducing risk of bias. Nevertheless bias would have occurred towards published as opposed to unpublished studies and English language manuscripts. Full data were not available for all studies, and although attempts were made to contact authors, these were unsuccessful. The clearly defined inclusion and exclusion criteria helped to correctly identify studies. However, by virtue of these selection criteria, only RCTs with sample sizes of great than 10 were included. Selection and quality assessment of studies was performed by only one reviewer, which represents a further source of potential bias.

4.5 Agreement/disagreement with other reviews

Consistent with Gregg et al. (2006), the current review endorses the finding that VRTs are equally efficacious as established psychological interventions for reducing anxiety disorder symptomatology. These results are also in line with the most recent – and methodologically robust – review (Meyerbroker et al., 2010), which concluded that currently there was no evidence to support the notion that VRTs are superior to traditional therapies for the majority of anxiety disorders. However, contrary to two previous meta-analyses (Powers & Emmelkamp, 2008; Parsons & Rizzo, 2008), the present review did not find VRTs to outperform traditional interventions in alleviating anxiety symptomatology. There are a number of potential explanations for this discrepancy.

First, this review exercised stringent methodological criteria for included studies, limiting results to RCTs with sample sizes of greater than ten. Conversely, in addition to RCTs Parsons and Rizzo (2008) included case reports and Powers and Emmelkamp (2008) reviewed uncontrolled trials; combining such disparate methodologies is questionable. Second, the current review analysed the outcomes of only one measure from each trial, whereas both previous meta-analyses aggregated results from all symptomatology outcomes, which may have produced different findings. Third, both previous meta-analyses mainly concerned specific phobias whereas the present review included a greater number of trials relating to social anxiety disorder, PTSD and panic disorder. It may be that feared stimuli in these more complex conditions are too idiosyncratic to rebuild in virtual environments and therefore better suited to traditional exposure methods (Meyerbroker et al., 2010). Indeed, examination of individual effect sizes shows that with the exception of one PTSD study (McLay et al., 2011), the most substantial effect sizes in favour of VRT came from three trials of specific phobias (Wiederhold et al., 2003; Tortella-Feliu et al., 2011; Emmelkamp et al., 2002). Similarly, with the exception of one aviophobia trial (Krijn et al., 2007) the three most sizable effects in

favour of traditional treatments came from studies of panic disorder and agoraphobia. It is therefore plausible that more complex conditions were less responsive to VRT, though continued controlled research is required to investigate this notion.

Consistent with previous reviews (e.g. Shahrbanian et al., 2009; Malloy & Milling, 2010), the present findings tentatively suggest that VRT shows promise in treating acute pain (experimental and procedural) but highlight a general paucity of controlled research, particularly for chronic conditions. Similarly, the findings from body-image distortion/obesity comparisons are in line with Gregg and Tarrier's (2007) conclusions that there is preliminary support for VRT as a comparatively efficacious treatment for obesity-related psychological outcomes. However, the results provide a less definitive and optimistic view of the literature than Riva et al. (2010) by highlighting the variable quality of trials and the lack of research in this domain. It is noted that Riva and colleagues have been responsible for most of the research into VRTs for eating disorders, which raises questions about the objectivity of their review.

4.6 Implications

4.6.1 Implications for practice

Anxiety disorders: Based on the current findings alone, it could be argued that there is no justifiable reason for introducing VRTs into routine clinical practice for anxiety disorders. VRT remains a relatively expensive alternative and does not appear to reduce symptomatology over-and-above traditional treatments, for which there is a strong existing evidence-base (Deacon & Abramowitz, 2004). However, the narrow focus of this review coupled with methodological imitations renders such conclusions premature. As aforementioned, the comparative efficacy of specific VRTs for specific anxiety disorders remains uncertain. For some conditions (e.g. specific phobias) VRTs may outperform traditional approaches both in terms of cost-

effectiveness (e.g. in aviophobia) and by way of exposing patients to scenarios that would be impossible to recreate in-vivo.

Conversely, it is possible that some anxiety-provoking stimuli may be too idiosyncratic to rebuild within the constraints of virtual environments and are therefore better suited to traditional approaches. However, as virtual environments become more readily amenable to therapeutic manipulation, there may be scope to recreate these highly idiosyncratic situations within the safety of a therapist's office. It is also possible that VRTs are better placed to target hard-to-reach patients such as those presenting with severe agoraphobia, unable to leave home to commence treatment. VRT offers the potential to introduce initial treatment sessions via home internet port-access (Meyerbroker et al., 2010), which may prove more cost-effective in the long-run. Conceptually this would fit well with existing self-administered interventions in the NHS, such as computerised CBT packages. However, continued research is required into each of these theoretical benefits before recommendations for practice can be made. Indeed, as highlighted by Cote and Bouchard (2008), VRTs were originally intended to address the shortcomings of traditional approaches rather than outperform them per se.

Pain: The present review provides tentative evidence that VRTs are superior to established methods of distraction in the management of acute and potentially chronic pain. Although preliminary, these findings suggest that VRT may offer exciting potential in routine practice, which could be particularly welcome in the remit of chronic pain. Unlike acute pain in which it is often possible to arrange in advance for a clinician to be present to deliver psychological interventions, chronic pain is largely reliant on self-management (Malloy & Milling, 2010). Consequently it is a particularly intractable, time-consuming and costly health problem. As the price of VR technology continues to fall, VR distraction may become an increasingly affordable and effective self-management tool for chronic pain patients. At present

however, a greater number of controlled trials are required in both acute and chronic pain before VRT initiatives can be trialled with confidence in clinical practice.

Body-image distortion/obesity: The present findings provide cautionary support for VRT as a comparable – possibly more effective – means of improving psychological outcomes related to obesity. The main theoretical advantages are the scope to manipulate a concept as subjective and hard to treat as body-image and provide patients with therapist-independent information about inaccurate internal representations of their body-image (Ferrer-García & Gutiérrez-Maldonado, 2012). This holds exciting potential not only in the clinical management of obesity outcomes but in anorexia and bulimia nervosa, for which body-image disturbance is implicated in the onset, maintenance and prognosis (Cash & Brown, 1987). These conditions are amongst the most prevalent in adolescents and young adults, and the current findings suggest that VRT could provide a means of targeting associated psychological outcomes. Future large-scale controlled trials are required to ascertain the effects of VRT on body-image distortion across the full spectrum of eating disorders. In particular, research is required beyond the work of Riva and colleagues, who have largely dominated this arena to date.

4.6.2 Implications for research

The current review highlights the need for greater number of large-scale and methodologically robust comparative efficacy trials with follow-up data, particularly in the domains of pain and obesity/body-image distortion. Thereafter separate meta-analyses are required for discrete psychological conditions in order to assess comparative efficacy of VRT for specific disorders. It will be important for future reviews to track changes in efficacy as VR equipment becomes more amenable to precise therapeutic manipulation and is increasingly trialled with complex clinical presentations and hard-to-access populations. Future research would benefit from

ensuring blinding of outcome assessors and gathering data on credibility and expectancy rates.

This review indicates a need for dismantling research in order to ascertain the relative contribution of different therapeutic components in VRT and assess its efficacy as a standalone treatment. Future studies may also benefit from including data on moderators such as 'presence'. If for example VR-exposure reduces fear by activating a fear structure (Foa & Kozak, 1986) then it is likely that the more 'real' VRT is, the more efficacious it will be (Powers & Emmelkamp., 2008). Given that presence is affected by a range of variables such as the sophistication of VR technology, future moderator analysis could inform decisions about whether to purchase immersive or non-immersive equipment and inform an understanding of the types of patients who may or may not be responsive to treatment.

Although symptomatology is an important outcome in establishing treatment efficacy, any decision to begin treatment would depend on a multitude of factors including patient preference, cost to the healthcare system, the impact of comorbidities and the severity of the disorder. Future research is required to understand how these factors can be considered alongside efficacy data in order to inform decisions about the utility of VRT in clinical practice. Data from behavioural avoidance tests, particularly in the anxiety disorders and obesity/body-image domains, also requires evaluation to better understand generalisability of effects to everyday life.

This review highlights some notable gaps in the literature such as the absence of comparative efficacy trials of VRT for anger management. Performance-based methods are a powerful means of producing therapeutic change across behavioral, cognitive, and affective modalities (Bandura, 1989) and accordingly CBT for anger typically includes role-play (Miyahira, Folen, Stetz & et al., 2010). VRTs have the potential to recreate ecologically valid anger-provoking scenarios in a safe

and ethical environment. Recent research indicates that aggressive virtual environments are an effective means of priming anger (Miyahira et al., 2010; Brinkman, Hattangadi, Meziane, & Pul, 2011). Therefore patients could gradually learn to manage arousal and negotiate prosocial responses with virtual characters. Developments in this arena could have far reaching consequences for the treatment of specific clinical conditions including oppositional defiance disorder, conduct disorder and antisocial personality disorder. This represents an exciting area for future development and comparative efficacy trials will be an important component of this research.

4.7 Summary

The current review suggests that VRTs are comparable – and in some instances more efficacious – than traditional interventions for reducing mental health symptomatology. However these conclusions are tentative in light of heterogeneity and methodological issues, coupled with limitations of the review process. There is a need for a greater number of methodologically robust large-scale studies, with follow-up data, particularly in the domains of pain and body-image distortion/obesity. Although this review provides a more sobering look at the literature than previous meta-analyses, outcomes other than symptomatology warrant further investigation in order to better understand the utility and cost-effectiveness of introducing VRTs into routine practice. There is also a need for research into treatment mediators and moderators; this would support the implementation of VRTs by making treatment outcomes more predictable. Finally, VRTs may offer exciting potential in new clinical applications, such as the full spectrum of eating disorders and in the treatment of anger.

REFERENCES

Astur, R.S., St. Germain, S.A., Baker, E.K., Calhoun, V.D., Pearlson, G.D., & Constable, R.T. (2005). fMRI Hippocampal Activity during a Virtual Radial Arm Maze. *Applied Psychophysiology and Biofeedback*, 30 (3), 307-17.

Bandura, A. (1989). Human Agency in Social Cognitive Theory. *American Psychologist*, 44, 1175–1184.

*Baños, R.M., Guillen, V., Quero, S., Garcia-Palacios, A., Raya, M.A., Botella, C. (2011). A virtual reality system for the treatment of stress-related disorders: A preliminary analysis of efficacy compared to a standard cognitive behavioral program. *International Journal of Human-Computing Studies*, 69(9), 602-613

Beck, A.T., Wright, F.D., Newman, C.F., & Liese, B.S. (1993). *Cognitive Therapy of Substance Abuse*. London: Guilford Press.

Blount, R.L., Piira, T., & Cohen, L.L. (2003). Management of paediatric pain and distress due to medical procedures. In Roberts, M.C. (Ed.), *Handbook of paediatric psychology* (pp. 216−233), 3rd ed. New York: Guilford.

Bohil, C.J., Alicea, B., & Biocca, F.A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, 12, 752-762.

Borenstein, M. (2005). Software for publication bias. In Rothstein, H.R. Sutton, A.J. & Borenstein, M. (Ed.), *Publication Bias in Meta-Analysis – Prevention,*

Assessment and Adjustments (pp. 193-220). London: John Wiley & Sons Ltd.

Botella, C., Garcia-Palacios, A., Baos, R., Quero, S., & Breton-Lopez, J. (2008). Virtual Reality in the treatment of pain. *Journal of CyberTherapy Rehabilitation*, 1 (1), 93-100.

*Botella, C., Villa, H., García-Palacios, A., Quero, S., Baños, R.M., Alcaniz, M., & Riva, G. (2007).Virtual reality exposure in the treatment of panic disorder and agoraphobia: A controlled study.*Clinical Psychology & Psychotherapy*, *14*, 164-175.

Brinkman, W. P., Hattangadi, N., Meziane, Z., & Pul, P. (2011). Design and evaluation of a virtual environment for the treatment of anger. *Proceedings of Virtual Reality International Conference* (VRIC 2011).

Bush, J. (2008). Viability of virtual reality exposure therapy as a treatment alternative. *Computers in Human Behavior,* 24 (3), 1032-1040.

Cash, T.F., & Brown, T.A. (1987). Body image in anorexia nervosa and bulimia nervosa: A review of the literature. *Behavior Modification*, 11, 487–521.

*Choi, Y.H., Vincelli, F., Riva, G., Wiederhold, B.K., Lee, J.H., and Park, K.H. (2005). Effects of group experiential cognitive therapy for the treatment of panic disorder with agoraphobia. *Cyberpsychology and Behavior*, *8*, 387-393.

Coelho, C.M., Waters, A.M., Hine, T.J., & Wallis, G. (2009). The use of virtual reality in acrophobia research and treatment. *Journal of Anxiety Disorders,* 23, 5, 563-574.

Clark, M.M., Abrams, D.B., Niaura, R S., Eaton, C. A., & Rossi, J S. (1991). Self-efficacy in weight management. *Journal of Consulting and Clinical Psychology*, 59, 739-744.

Carlin, A.S., Hoffman, H.G., & Weghorst, S. (1997). Virtual reality and tactile augmentation in the treatment of spider phobia: A case report. *Behaviour Research and Therapy*, *35*, 153-158.

Cohn, L.D., & Becker, B.J. (2003). How meta-analysis increases statistical power. *Psychological Methods, 8*, 243–253.

Cooper, Z., Fairburn, C.G., Hawker, D. (2003). *Cognitive Behavioral Treatment of Obesity: A Clinician's Guide*. New York: Guilford Press.

Côté, S. & Bouchard, S. (2008). Virtual reality exposure for phobias: A critical review. *Journal of CyberTherapy and Rehabilitation*, 1(1), 75-91.

Crowther, J.H., Sanftner, J., Bonifazi, D.Z., & Shepard, K.L. (2001). The role of daily hassles in binge eating. *International Journal of Eating Disorders*, *29*, 449–454.

Da Costa, R.T., Sardinha, A., & Nardi, A.E. (2008). Virtual reality exposure in the treatment of fear of flying. *Aviation, Space & Environmental Medicine, 79* (9), 899-903.

Da Costa, R.T., Carvalho, M.R., & Nardi, A.E. (2010). Virtual reality exposure therapy in the treatment of driving phobia: A systematic review. Psicologia: Teoria e Pesquisa, 26 (1), 131-137.

Deacon, B.J., & Abramowitz, J.S. (2004). Cognitive and behavioral treatments for anxiety disorders: A review of meta-analytic findings. *Journal of Clinical Psychology, 60,* 429-441.

Difede, J., Hoffman, H. (2002). Virtual reality exposure therapy for PTSD following the WTC: A case report. *Cyberpsychology & Behavior*, 5(6), 529-535.

DiNardo, P.A., Brown, T.A., & Barlow, D.H. (1994). *Anxiety Disorders Interview Schedule for DSM-IV: Lifetime Version (ADIS-IV-L).* New York: Graywind Publications.

Downs, S.H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomized studies of health care interventions. *Journal of Epidemiology Community Health, 52*, 377-384.

Ehrsson, H.H. (2007). The experimental induction of out-of-body experiences. *Science* 317, 1048.

Emmelkamp, P.M.G., Bruynzeel, M., Drost, L., & van der Mast, C.A. (2001). Virtual reality treatment in acrophobia: a comparison with exposure in vivo. *Cyberpsychology and Behavior*, *4*, 335–340.

*Emmelkamp, P .M.G., Krijn, M., Hulsbosch, A., de Vries, S., Schuemie, M.J., & van der Mast, C.A.P.G. (2002).Virtual reality treatment versus exposure in vivo: a comparative evaluation in acrophobia. *Behaviour Research and Therapy*, *40*, 509–516.

Ferrer-García, M., Gutiérrez-Maldonado, J. (2012). The use of virtual reality in the study, assessment, and treatment of body image in eating disorders and nonclinical samples: a review of the literature. *Body Image*, 9(1), 1-11.

First, M.B., Spitzer, R.L., Williams, J.B.W., Gibbon, M. (1995): Structured Clinical Interview for DSM-IV (SCID-I) - Research Version, *Bometrics Research*, New York: NY Press.

Flynn, D., Schaik, P.V., Blackman, T., Hobbs, B., Calderdon, C. (2003). Developing a virtual reality based methodology for people with Dementia: a feasibility study. *CyberPsychol Behav,* 6, 591–611.

Foa, E.B., & Kozak, M.J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin*, 99(1), 20–35.

Friedman, M.A., & Brownell, K.D. (1995). Psychological Correlates of Obesity: Moving to the Next Research Generation.” *Psychological Bulletin,* 117, 3–20.

Geliebter, A., & Aversa, A. (2003). Emotional eating in overweight, normal weight, and underweight individuals. *Eating Behaviour*, 3, 341–7.

*Gershon, J., Zimand, E., Pickering, M., Rothbaum, B.O., & Hodges, L. (2004). A pilot and feasibility study of virtual reality as a distraction for children with cancer. *Journal of the American Academy of Child Adolescent Psychiatry, 43(10)*, 1243–1249.

Glantz, K., Rizzo, A., & Graap, K. (2003). Virtual reality for psychotherapy: Current reality and future possibilities. *Psychotherapy: Theory, Research, Practice, Training, 40,* 55-67.

Gorini, A., Mosso, J.L., Mosso, D., Pineda, E., Ruíz, N.L., Ramíez, M., et al. (2009). Emotional response to virtual reality exposure across different cultures: the role of the attribution process. *CyberPsychology & Behavior*, 12(6), 699–705.

Gregg, L., & Tarrier, N. (2007). Virtual reality in mental health: a review of the literature. *Social Psychiatry Psychiatric Epidemiology*, 42 (5), 343-354.

Griffin, M.J. (1990). *Handbook of human vibration*. London: Academic Press.

Higgins, J.P.T., Green, S. (2005). *Cochrane Handbook for Systematic Reviews of Interventions*. www.cochrane.org/resouces/handbook/hbook.htm 4.2.5

Hoffman, H.G., Seibel, E J., Richards, T.L., Furness, T.A., Patterson, D.R., & Sharar, S.R. (2006). Virtual reality helmet display influences the magnitude of virtual reality analgesia. *Journal of Pain*, 7, 843−850.

Kalawsky, R.S. (1993). *The science of virtual reality and virtual environments. A technical, scientific and engineering reference on virtual environments.* Wokingham: Addison-Wesley Publishing.

Klinger, E., Bouchard, S., Légeron, P., Roy, S., Lauer, F., Chemin, I., & Nugues, P. (2005). Virtual reality therapy versus cognitive behavior therapy for social phobia: a preliminary controlled study.*CyberPsychology & Behavior*, 8(1), 76-88.

Krijn, M., Emmelkamp, P.M., Olafsson, R.P., & Biemond, R. (2004). Virtual reality exposure therapy of anxiety disorders: A review. *Clinical Psychology Review*, 24, 259–281.

Krijn, M., Emmelkamp, P.M.G., Biemond, R., de Wilde de Ligny, C., Schuemie, M. J., & van der Mast, C.A.P.G. (2004). Treatment of acrophobia in virtual reality: The role of immersion and presence. *Behaviour Research and Therapy, 42*, 229- 239.

*Krijn, M., Emmelkamp, P.M.G., Ólafsson, R.P., Bouwman, M., van Gerwen, L., Spinhoven, P., Schuemie, M.J. & van der Mast, C.A.P.G. (2007). Fear of flying treatment methods: virtual reality exposure vs. cognitive behavioral therapy. *Aviation, Space and Environmental Medicine,* 78(2),121-128.

Ku, J.H., Cho, W., Kim, J., Peled, A., Wiederhold, B.K., Wiederhold, M.D., Kim, I.Y., Lee, J.H., & Kim, S.I. (2003). A virtual environment for investigating schizophrenic patients' characteristics: assessment of cognitive and navigation ability. *CyberPsychol Behav,* 6, 397–404.

Lee, J.H., Ku, J.H., Cho, W., Hahn, W.Y., Kim, I.Y., Lee, S., Kang, Y., Deog, Y.K., Yu, T., Wiederhold, B.K., Wiederhold, M.D., & Kim, S.I. (2003) A virtual reality system for the assessment and rehabilitation of the activities of daily living. *CyberPsychol Behav,* 6, 383–388.

*Leibovici, V., Magora, F., Cohen, S., & Ingber, A. (2009). Effects of virtual reality immersion and audiovisual distraction techniques for patients with pruritus. *Pain Research and Management*, 14, 283−286.

Mahrer, N.E., & Gold, J.I. (2009). The use of virtual reality for pain control: a review. *Current Pain & Headache Reports*, 13, 100-109.

Malloy, K.M., & Milling, L.S. (2010). The effectiveness of virtual reality distraction for pain reduction: A systematic review. *Clinical Psychology Review*, 30, 1011-1018.

Manzoni, G.M., Gorini, A., Preziosa, A., Pagnini, F., Castelnuovo, G., Molinari, E., & Riva, G. (2008). New Technologies and Relaxation: An Explorative Study on Obese Patients with Emotional Eating. Journal of CyberTherapy and Rehabilitation,* 1(2), 182–192.

Marlatt, G.A., & Gordon, J.R. (1985). *Relapse Prevention: Maintenance strategies in the treatment of addictive behaviours*. New York: Guilford Press.

McCaul, K.D., & Malott, J.M. (1984). Distraction and coping with pain. *Psychological Bulletin*, 95, 516−533.

*McLay, R.N., Wood, D.P., Webb-Murphy, J.A., Spira, J.L., Wiederhold, M.D., Pyne, J.M., & Wiederhold. B.K. (2011). A Randomized, Controlled Trial of Virtual Reality-Graded Exposure Therapy for Post-Traumatic Stress Disorder in Active Duty Service Members with Combat-Related Post-Traumatic Stress Disorder. *Cyberpsychology, Behavior, and Social Networking*, 14(4), 223-229.

Meyerbroker. K.., & Emmelkamp, P.M.G. (2010). Virtual Reality Exposure Therapy in Anxiety Disorders: A Systematic Review of Process-and-Outcome Studies. *Depression and Anxiety* 27, 933-944.

*Michaliszyn, D., Marchand, A., Bouchard, S., Martel, M.O., Poirier-Bisson, J. (2010). A randomized, Controlled Clinical Trial of In Virtuo and In Vivo Exposure for Spider Phobia. *Cyberpsychology, behavior and Social Networking*, 13(6), 689-695.

Miyahira, S.D., Folen,R.A., Stetz, M., Rizzo, A., & Kawasaki, M.M. (2010). Use of immersive virtual reality for treating anger. *Studies in health technology and informatics*, 154, 82-6

*Mott, J., Bucolo, S., Cuttle, L., Hilder, M., Miller, K., & Kimble, R.M. (2008). The efficacy of an augmented virtual reality system to alleviate pain in children undergoing burns dressing changes: A randomised controlled trial. *Burns*, 34, 803−808.

*Mühlberger, A., Wiedemann, G., & Pauli, P. (2003). Efficacy of a one-session virtual reality exposure treatment for fear of flying. *Psychotherapy Research, 13*, 323-336.

Nilsson, S., Finnstrom, B., Kokinsky, E., & Enskar, K. (2009). The use of Virtual Reality for needle-related procedural pain and distress in children and adolescents in a paediatric oncology unit. *European Journal of Oncology Nursing*, 13, 102−109.

Norcross, J.C., Hedges, M., & Prochaska, J.O. (2002). The face of 2010: A Delphi poll on the future of psychotherapy. *Professional Psychology: Research & Practice*, 33, 316-322.

Ong, I., Linden, W.,& Young, S. (2004) Stress management what is it? *Journal of Psychosomatic Research*, 56, 133-137.

Parsons, T.D., & Rizzo, A.A. (2008). Affective Outcomes of Virtual Reality Exposure Therapy for Anxiety and Specific Phobias: A Meta-Analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 39,* 250-261.

*Patterson, D.R., Wiechman, S.A., Jenson, M., & Sharar, S.R. (2006). Hypnosis delivered through immersive virtual reality for burn pain: a clinical case series. *International Journal of Clinical and Experimental Hypnosis*, 54 (2), 130-142.

Penate, W., Pitti, C.T., Bethencourt, J.M., de la Fuente, J., & Gracia, R. (2007). The effects of a treatment based on the use of virtual reality exposure and cognitive-behavioral therapy applied to patients with agoraphobia. *International Journal of Clinical and Health Psychology*, 8(1), 5-22.

*Pérez-Ara, M.A., Quero, S., Botella, C., Baños, R., Andreu-Mateu, S., García-Palacios, A., & Bretón-López, J. (2010). Virtual reality interoceptive exposure for the treatment of panic disorder and agoraphobia. *Annual Review of Cybertherapy and Telemedicine*, 8, 61-64.

Powers, M.B., & Emmelkamp, P.M. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders*, 22(3), 561-569.

Price, M., Anderson, P., & Rothbaum, B.O. (2008). Virtual reality as treatment for fear of flying: A review of recent research. *International Journal of Behavioral Consultation and Therapy*, 4(4), 309–315.

Pull, C.B. (Ed.). (2005). Current status of virtual reality exposure therapy in anxiety disorders: Editorial review. *Current Opinion in Psychiatry*, 18, 7–14.

*Riva, G., Bacchetta, M., Baruffi, M., & Molinari E. (2001). Virtual Reality–Based Multidimensional Therapy for the Treatment of Body Image Disturbances in Obesity: A Controlled Study. *CyberPsychology & Behavior*, 4, 511-527.

Riva, G., & Melis, L. (1997). Virtual reality for the treatment of body image disturbances. In Riva, G. (Ed.), *Virtual reality in neuro-psycho-physiology: cognitive, clinical and methodological issues in assessment and rehabilitation,* pp. 95–111. Amsterdam: IOS Press.

*Riva, G., Bacchetta, M., Baruffi, M., Molinari, E. (2002). Virtual reality–based multidimensional therapy for the treatment of body image disturbances in binge eating disorders: a preliminary controlled study. *IEEE Transactions on Information Technology in Biomedicine*, 6, 224–234.

*Riva, G., Bacchetta, M., Cesa, G., Conti, S., Castelnuovo, G., Mantovani, F., & Molinari, M. (2006). Is severe obesity a form of addiction? Rationale, clinical approach, and controller clinical trial. *CyberPsychology and Behaviour*, 9, 4, 457-479.

Riva, G., Molinari, E. (2009). Virtual Reality in Eating Disorders and Obesity. *CyberTherapy & Rehabilitation*, 2(2), 16-19.

Rizzo, A.A., & Kim, G.J. (2005). A SWOT analysis of the field of VR rehabilitation and therapy, *Presence: Teleoperators and Virtual Environments* 14 (2), 119-146.

Roberts, N.P., Kitchiner, N.J., Kenardy, J., & Bisson, J.I. (2010). Multiple session early psychological interventions for the prevention of post-traumatic stress disorder. *Cochrane Database of Systematic Reviews*, 8 (3).

*Robillard, G., Bouchard, S., Dumoulin, S., Guitard, T., & Klinger, E. (2010). Using Virtual Humans to Alleviate Social Anxiety: Preliminary Report from a

Comparative Outcome Study. *Annual Review of Psychotherapy and Telemedicine*, 8, 46-49.

*Rothbaum, B.O., Zimand, E., Hodges, L., Lang, D., & Wilson, J. (2006). Virtual reality exposure therapy and standard in vivo exposure therapy in the treatment of fear of flying. *Behaviour Therapy*, 37, 80–90.

Rothbaum, B.O., Hodges, L., Smith, S., Lee, J.H., & Price, L. (2000). A controlled study of virtual reality exposure therapy for the fear of flying. *Journal of Consulting and Clinical Psychology*, 68, 1020–1026.

Rothbaum, B.O., Hodges, L.F., Ready, D., Graap, K., & Alarcon, R. D. (2001). Virtual reality exposure therapy for Vietnam veterans with posttraumatic stress disorder. *Journal of Clinical Psychiatry*, 62, 617–622.

Sadowsky, J., & Massof, R.W. (1994). Sensory engineering: the science of synthetic environments. *John Hopkins APL Technical Digest*, 15, 99–109.

Shahrbanian, S., Ma, X., Korner-Bitensky, N., Simmonds, M.J. (2009). Scientific evidence for the effectiveness of virtual reality for pain reduction in adults with acute or chronic pain. *Studies in health technology and Informatics*, 144, 40-3.

Sheehan, D., Lecrubier, Y., Harnett-Sheehan, K., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. (1998). The Mini International Neuropsychiatric Interview (M.I.N.I.): the development & validation of a

structured diagnostic psychiatric interview. *Journal of Clinical Psychiatry, 59*, 22-33.

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., & Sanchez-Vives, M.A. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PLoS ONE,* 1, e39.

Spielberger, C.D., Gorsuch, R.L., & Lushene, R.E. (1970). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

St-Jacques, J., Bouchard, S., & Bélanger, C. (2010). Is virtual reality effective to motivate and raise interest in phobic children towards therapy? *Journal of Clinical Psychiatry*, 71(7), 924-931.

Tarrier, N., Pilgrim, H., Sommerfield, C., Faragher, B., Reynolds, M., Graham, E.A. (1999). A randomized trial of cognitive therapy and imaginal exposure in the treatment of chronic post-traumatic stress disorder. *Journal of Consulting and Clinical Psychology*, 67, 13–18.

Thompson, J.K., Heinberg, L.J., Altabe, M., & Tantleff-Dunn, S. (1999). *Exacting beauty: theory, assessment and treatment of body image disturbance*. Washington, DC: American Psychological Association.

*Tortella-Feliu, M., Botella, C., Llabrés, J., Bretón-López, J.M., del Amo, A.R., Baños, R.M., & Gelabert, J.M.(2011). Virtual reality versus computer-aided exposure treatments for fear of flying. *Behaviour Modification*, 35(1), 3-30.

Turk, D.C. (2002). Clinical effectiveness and cost-effectiveness of treatments for patients with chronic pain. *Clinical Journal of Pain*, 18, 355−365.

Vincelli, F. (1999). From Imagination to Virtual Reality: The Future of Clinical Psychology. *Cyberpsychology and Behaviour,* 2 (3), 241-248.

*Wallach, H.S., Safir, M.P., & Bar-Zvi, M. (2009). Virtual reality cognitive behavior therapy for public speaking anxiety; a randomized clinical trial. *Behaviour Modification*, 33, 314–338.

*Wiederhold, B.K., Jang, D.P., Gevirtz, R.G., Kim, S.I., Kim, I.Y., & Wiederhold, M.D. (2002). The treatment of fear of flying: A controlled study of imaginal and virtual reality graded exposure therapy. *IEEE Transactions on Information Technology in Biomedicine*, 6(3), 218–223.

Wismeijer, A.A.J.**,** & Vingerhoets, A.J.J M. (2005). The use of Virtual Reality and audiovisual eyeglass systems as adjunct analgesic techniques: A review of the literature. *Annals of Behavioral Medicine, 30,* 268-278.

Witmer, B.G., & Singer, M.J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire, *Presence: Teleoperators and Virtual Environments*, 7(3), 225-240.

Part 2: Empirical Paper

Emotional reactivity and aggression in antisocial personality disorder: a virtual reality study

ABSTRACT

Background: Robust differences have been observed between aggression-prone and non-aggression-prone groups in their emotional and behavioural reactivity during stress and provocation. However, there is a lack of ecologically valid research, particularly with regard to individuals with antisocial personality disorder (ASPD).

Aims: To investigate whether individuals diagnosed with ASPD, relative to healthy volunteers, show increased negative emotional reactivity and more aggressive behaviour during virtual-reality (VR) provocation.

Method: Fifteen individuals diagnosed with ASPD and 20 healthy volunteers witnessed aggression escalating between characters in a virtual 'pub'. Emotional reactivity was indexed across affective, cognitive and behavioural domains using questionnaires, interview and observation.

Results: Following VR provocation, participants showed significant elevations in negative affect but no changes in positive affect. Clinical participants displayed greater increases in negative affect, more anger-oriented cognitions, less victim-empathic cognitions, less conciliatory/prosocial behaviour, and a trend towards more aggressive behaviour than healthy volunteers.

Conclusions: It is tentatively concluded that participants with ASPD experienced more marked negative emotionality in response to VR provocation than healthy volunteers, contributing to a reduction in prosocial behaviour and a trend towards more impulsive aggression. Potential clinical moderators and methodological factors are considered as explanations for aggressive behavioural differences failing to meet significance. Implications for how ASPD aggression is conceptualised, which interventions may be indicated and how to assess and monitor risk are discussed. Findings are exploratory and methodologically robust modified replications are required before firm conclusions can be reached about the role that emotional reactivity plays in aggression within this population.

"Anybody can become angry – that is easy, but to be angry with the right person, and to the right degree, and at the right time, and for the right purpose, and in the right way – that is not within everybody's power and is not easy": Aristotle

1. INTRODUCTION

1.1 Antisocial Personality Disorder

Antisocial personality disorder (ASPD) is characterised by "a pervasive pattern of disregard for, and violation of the rights of others" (Diagnostic and Statistical Manual of Mental Disorders, revised 4th edition [DSM-IV]; American Psychiatric Association [APA], 2000). In the UK, the prevalence of ASPD amongst men and women is estimated at 1% and 0.2%, respectively (Coid, Tyer & Roberts, 2006). Comorbidity studies suggest that 90.4% of people who meet criteria for ASPD have at least one other psychiatric condition (Swanson, Bland & Newman, 1994), most commonly substance misuse (Robins & Price, 1991).

ASPD is associated with profound interpersonal and social disadvantage across the lifespan, including low educational attainment, imprisonment and premature mortality (Swanson et al., 1994). Thus whilst criminality is central to the definition of ASPD (DSM IV; APA, 2000), it is often the consequence of long-standing difficulties and is not the only feature. Consequently, ASPD is conceptualised as distinct from criminal behaviour/violence per se (NICE CG 77, 2009), evidenced by the fact that roughly 50% of prisoners do not meet diagnostic criteria (e.g. Fazel & Danesh, 2002). A further distinction to be made is between ASPD and psychopathy. Psychopathy is characterised by a cluster of interpersonal/affective features (e.g. callous/lacking empathy) and marked social deviance (e.g. parasitic lifestyle) (Hare, 2003). Whilst the two are associated, only 10% of those with ASPD fulfil criteria for psychopathy (NICE CG 77, 2010).

1.2 Emotional processing in aggressive populations

Research into the processes underlying aggression and violence in criminal populations has focused almost exclusively on psychopathy, in which chronic physiological under-arousal is a well-replicated correlate of stress (e.g. Patrick, Cuthbert & Lang, 1994). By implication, aggression/violence committed by this subgroup is considered primarily 'instrumental' and premeditated (e.g. Crick & Dodge, 1996), without anger necessarily playing a crucial role (Porter, Woodworth, Earle, Drugge & Boer, 2003). However, this cannot be assumed applicable to ASPD, which by definition entails deregulated and uncontrollable anger (APA, 2000). Indeed, research suggests that whilst psychopathic offenders show reduced behavioral and physiological responses to emotional stimuli (Levenston, Patrick, Bradley, & Lang, 2000) such deficits do not emerge for individuals diagnosed with ASPD (Vaidyanathan, Hall, Patrick, & Bernat, 2011). This implies that emotional hyper-responsivity may play a more pivotal role in ASPD aggression, by way of impulsive reactions to provocation or stress.

In support of this, brain-imaging studies of violent criminals show dysfunction in the frontocortical (e.g. Blake, Pincus & Buckner, 1995) and limbic regions (e.g. Raine, Buchsbaum & LaCasse, 1997) that mediate emotional processing. Research also demonstrates that the impulsive traits of ASPD are associated with negative emotional reactivity (Hicks & Patrick, 2006), and with deficits in brain regions that govern inhibitory control (Morgan & Lilienfeld, 2000). Further evidence indicates that aggressive children show autonomic acceleration during stress (Lorber, 2004) and that weak vagal control is associated with externalising problems in at-risk children (e.g. El-Sheikh, Harger & Whitson, 2001).

Experimental research also implicates emotional hyper-responsivity in aggressive behaviour. Verona, Patrick & Lang (2002) subtyped healthy volunteers into those scoring 'high' and 'low' trait negative emotionality and induced negative

affect via a series of aversive air blasts. Participants who scored high on trait negative emotionality exhibited enhanced tonic distress and delivered more intense electric shocks to an experimental confederate during air blast intervals than low scorers. This suggested that sustained negative affect biases high stress-reactive individuals towards more intense and prolonged acts of aggression. Similarly, correlational research indicates that high levels of trait negative emotionality are characteristic of domestic assaulters (Farrington, 1980) and partner abuse is more likely to occur in times of stress (Rosenbaum & O'Leary, 1981).

These findings can be understood in terms of Berkowitz's (1990) cognitive-neoassociationist model of aggression, which conceptualises emotional states as a collective network in which affect, cognitions, physiological reactions and motor responses are all linked together. Because these interconnected components involve adaptive mobilisation for defensive action (Lang, Bradley, & Cuthbert, 1990), negative affect primes aggressive behaviour during provocation or stress (Berkowitz, 1983, 1994). By implication, individuals who are vulnerable to emotional hyper-responsivity are particularly likely to display aggression under provocative conditions.

1.2.1 Emotional reactivity in ASPD

The aforementioned findings raise the possibility that unlike psychopathy, ASPD entails difficulties regulating anger (and other emotional reactions), with consequential enhancement of defensive reactivity under conditions of perceived threat. However, only one study to date is known to have directly investigated emotional responsivity in relation to ASPD. Lobbestael, Arntz, Cima & Chakhssi (2009) compared ASPD participants to healthy volunteers and a borderline personality disorder group on indices of emotional reactivity at baseline and following an anger-induction interview. Compared with control groups, ASPD participants displayed comparative cognitive hyper-responsivity and physiological

hypo-responsivity but no differences in self-reported anger. These findings were understood in two ways. First, because lying is a central characteristic of ASPD, questionnaire data may have reflected 'text appropriate' ratings (Herpertz, Werth, Lukas et al., 2001), thereby artificially reducing ASPD anger scores. Second, the discrepancy between cognitive hyper-responsivity and physiological hypo-responsivity within the ASPD group was seen to reflect animal defences, in which the 'fight phase' is preceded by an 'orienting phase' (e.g. Lacey & Lacey, 1958). Because orientation is characterised by cognitive vigilance and decelerated heart rate (Lang, David & Ohman, 2000), it was tentatively concluded that ASPD participants remained in this phase longer than control participants. This was attributed to the anger-induction interview representing a 'remote threat' (e.g. the object of aggression was far away and therefore not set up to provoke a counter-attack) that was insufficient to provoke a fight response in ASPD participants.

Lobbestael et al. (2009) recommended that future research utilise imminent/intense anger-induction methods, in order to establish whether ASPD participants switch from physiological hypo-responsivity into an extreme defensive response, resulting in impulsive aggression/violence. However until recently the scope of experimental anger induction has been limited, reflecting the tension between ethics, safety and generalisability (Ferguson & Rueda, 2009). Virtual reality (VR) offers a novel potential means of balancing these priorities within a controlled laboratory environment.

1.3 Virtual reality

An immersive virtual environment is a computer-generated real-time surrounding that displays virtual sensory data from a viewpoint determined by the tracked position and orientation of the occupant's head (Rizzo & Kim, 2005). This delivers a life-sized VR within which a person can experience events and interact with virtual objects and characters (avatars). Research indicates that people

respond realistically to virtual environments (Rovira, Swapp, Spanlang & Slater, 2009), even in extreme social situations (Slater, Antley, Davison, et al., 2006). This has been capitalised on in the treatment of a range of psychological disorders (e.g. Muhlberger, Wiedermann & Pauli, 2003) and in investigations into the processes underlying these conditions (for a detailed review see Bohil, Alicea & Boicca, 2012).

There have been a number of recent developments in the field of VR-simulated violence and aggression. Rovira and colleagues (2009) developed a virtual 'pub' to study participants' responses to a violent dispute between two avatars. An exploratory study indicated that most participants attempted to physically or verbally intervene in the confrontation, despite awareness that the avatars were not real. Two recent studies have directly substantiated the anger-arousal potential of VR using physiological measurement (Brinkman, Hattangadi, Mezinae & Pul, 2011), behavioural observations and self-report (Miyahira, Folen, Hoffman, Garcia-Palacios & Schaper, 2010). Together these findings indicate that VR provides an ecologically valid means of studying how people respond to provocation, enabling researchers to observe complex behaviour and decision-making, rather than being restricted to more arbitrary measures.

In summary, there is some evidence to suggest that ASPD aggression may reflect emotional hyper-responsivity but previous research has focused almost exclusively on psychopathic and aggression-prone individuals. The only study known to have investigated anger induction in ASPD (Lobbestael et al., 2009) did not measure behaviour, meaning that the link between emotional reactivity and actual aggression remains unclear. Moreover, the stress-induction interview used was a remote threat unlikely to typify the real-life triggers of reactive anger, which leaves outstanding questions about the emotional and behavioural reactivity of individuals with ASPD during ecologically valid conditions of imminent threat. VR offers a method for addressing these gaps in the literature. Improved understanding

could have implications for how ASPD aggression is conceptualised, which interventions are best indicated and how to assess and monitor risk.

1.4 The current research

The present study aimed to investigate differences between individuals with a diagnosis of ASPD and healthy volunteers in their emotional and behavioural responses to experimentally induced provocation. Specifically it aimed to establish whether the former displayed more negative affect, anger-oriented cognitions and aggressive behaviour when exposed to violence in a VR pub scenario (Rovira et al., 2009). The original research intention was to index the full associative network comprising emotions (Berkowitz, 1990), including autonomic arousal (via electrocardiogram) and motor activity (via electromyogram). However, unfortunately both strands of physiological data proved uninterpretable due to an equipment failure that was not detected until completion of testing. Consequently, the final sources of data comprised (i) affective states measured by the Positive and Negative Affect schedule (PANAS; Watson, Clark & Tellegen, 1988), (ii) cognitive appraisals/interpretations elicited via semi-structured interview, and (iii) behavioural reactions during VR.

1.4.1 Hypotheses

The hypotheses were informed by previous research demonstrating differences between aggression-prone and non-aggression-prone groups in their emotional reactivity during provocation (Lobbestael et al., 2010), and their propensity to behave aggressively when exposed to stress (Verona et al., 2002). Accordingly it was predicted that participants with ASPD, relative to healthy volunteers, would display:

1. *Affect*: A greater increase on the negative scale, and a greater decrease on the positive scale of the PANAS between baseline and post-VR.
2. *Cognitions*: More anger-oriented and less victim-empathic cognitions during interview.
3. *Behaviour*: More aggressive/antagonistic behaviour and less conciliatory/prosocial behaviour during VR.

In addition to the foregoing, the relationship between emotional reactivity and behaviour was investigated based on the following subsidiary hypothesis:

4. *Relationship between outcomes*: Regardless of group membership, antagonistic behaviour will correlate with negative affect and anger-oriented cognitions, and conciliatory behaviour will correlate with positive affect and victim-empathic cognitions.

2. METHOD

2.1 Design

This research utilised a 2x2 group-comparison mixed design. The between-groups variable was 'clinical group' (ASPD vs. healthy volunteers) and the repeated-measures variable 'point of measurement' (baseline vs. post-VR). Dependent variables were 'affective states', 'cognitions' and 'VR intervention behaviours'. All participants took part in the same VR scenario and completed self-report questionnaires and an interview.

2.2 Participants

Clinical participants were 15 males, aged 24-62 years (M = 44). All had an established primary diagnosis of ASPD (DSM-IV; APA, 2000) or Dissocial Personality Disorder (ICD-10; WHO, 1992), previously diagnosed by a psychiatrist

using the Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II; First, Williams, Spitzer & Gibbon, 1997) and/or the International Personality Disorder Examination (IPDE; Loranger, Sartorius, Andreoli, Berger et al., 1994). Clinical participants were recruited from three specialist forensic personality disorder services (inpatients *n* = 4, community patients n = 9), one generic forensic service (community patient *n* = 1) and one non-forensic personality disorder service (community patient *n* = 1). Inclusion criteria were: (1) a well-defined psychiatric diagnosis of ASPD, (2) one or more conviction(s) for a criminal offence, and (3) capacity to consent to the research. Inpatients detained under the Mental Health Act (1983) also required community leave (agreed by the Ministry of Justice and respective clinical team) of a sufficient allowance to participate in the study. Exclusion criteria were active symptoms of major mental illness and/or a diagnosed learning disability.

Healthy volunteers were 20 male students of University College London (UCL) and members of the wider London community aged 20-47 years (*M* = 28), recruited from UCL participant panels and via advertisements placed around campus (Appendix 2). Exclusion criteria were: a psychiatric and/or violent offending history.

General inclusion criteria for clinical and non-clinical groups were: male participants, aged 18-65, with sufficient command of the English language to understand the VR scenario and questionnaires. Participants with epilepsy were excluded because VR can trigger seizures. Demographic and clinical characteristics of participants are displayed in Table 1 and Table 2 in the Results section.

2.3 Sample size and statistical power

Power analysis for this study was informed by prior work by Lobbestael et al. (2009). In their study an anger-induction interview was used to compare individuals

with ASPD and control groups on various indices of emotional reactivity. On a measure of self-reported anger, ASPD participants compared to healthy volunteers showed a between-groups effect size of d = 0.53 (Cohen's d = medium). Power calculations for a repeated-measures, between-factors ANOVA were conducted in G*Power 3.1.2 (Faul, Erdfelder, Lang and Buchner, 2007), specifying $\alpha = 0.05$ and desired power at 0.80. For a comparison between two groups, with two levels of measurement, the required sample size was 24 (12 per group). The achieved sample size was 35 (15 ASPD and 20 healthy volunteers).

2.4 Procedure

2.4.1 Participant identification and recruitment

The project was approved by Camberwell & St Giles NHS Research Ethics Committee (Appendix 3) and UCL Ethics Committee (Appendix 3). To recruit clinical participants, the researcher circulated a letter (Appendix 2) to clinicians in respective services requesting referrals for the research. Potential participants' medical records were screened by a member of their clinical team, who approached eligible candidates with the information sheet (Appendix 4) to ascertain whether they were interested in taking part.

The names of interested participants were passed to the researcher, who sought permission from the Responsible Clinician/Consultant Psychiatrist to meet with the participant directly. The purpose of this initial meeting was to describe the research in full, explain the process of informed consent and provide an opportunity for questions. For inpatients this introduction took place on hospital premises, facilitated by a member of their clinical team. Initial introductions with community patients occurred either at their respective hostels or by telephone. Thereafter research appointments were scheduled with participants to attend UCL and the Responsible Clinician/Consultant Psychiatrist was informed of their involvement

(Appendix 2). Healthy volunteers were a self-selecting sample recruited via participant panels and advertisements at UCL. Participants were asked to contact the researcher directly, whereupon they were sent a copy of the information sheet (Appendix 4) and a research appointment was scheduled. A TREND diagram (Des Jarlais, Lyles & Crepaz, 2006) of non-randomised recruitment is shown in Figure 1.

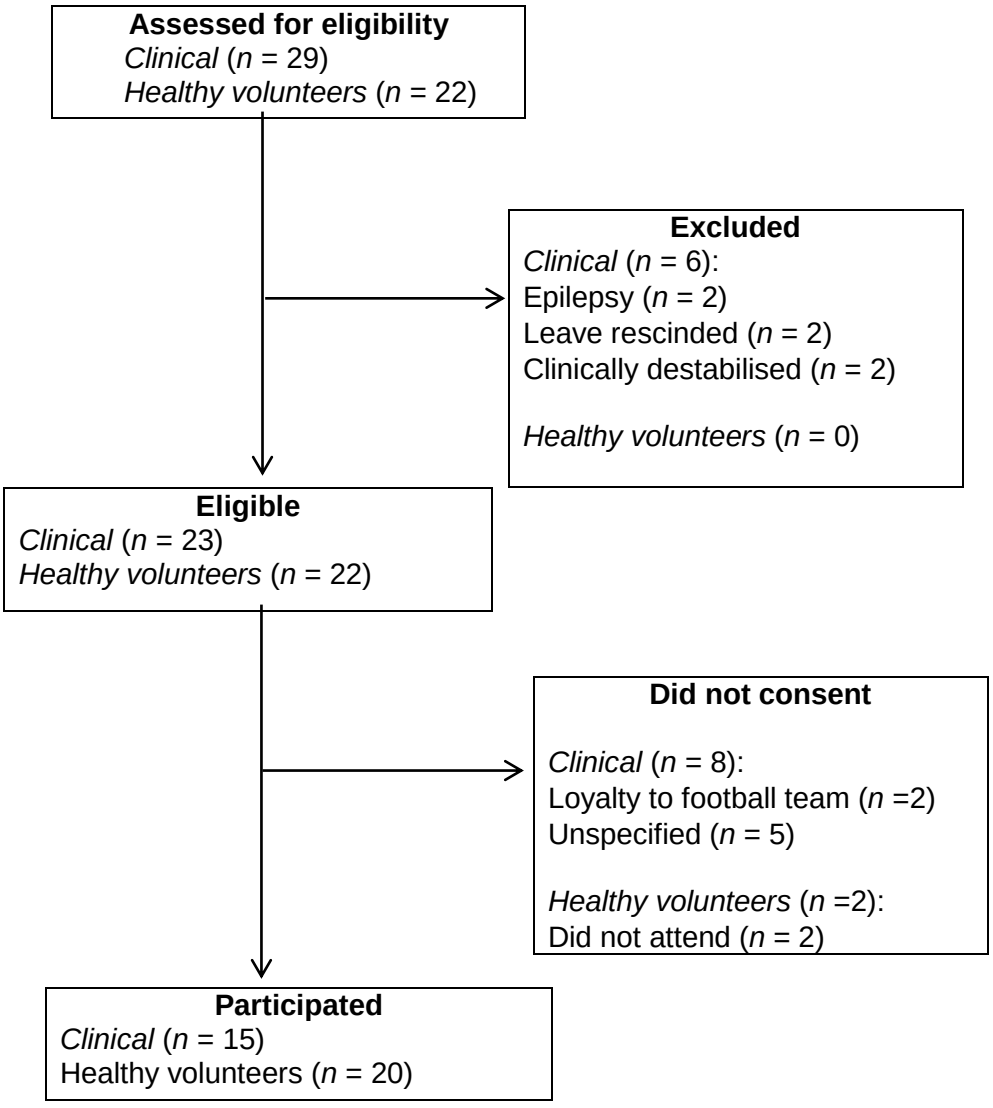


Figure 1: TREND diagram of the non-randomised recruitment process

2.4.2 Testing procedure

Testing took place in the Department of Computer Science, UCL where the VR suite was situated. Written informed consent (Appendix 4) and basic

demographic information were obtained from all participants prior to testing. Testing took 60-90 minutes and occurred in three main stages. Prior to entering VR, participants completed the baseline PANAS; physiological measures were also obtained at this stage. Stage two involved entering the VR environment. There was an initial three-minute period in which participants became accustomed to the VR pub and familiarised themselves with the equipment. Participants then took part in the experimental environment, which depicted escalating violence between an aggressor and a victim in the pub. Participants were asked to wear an Arsenal scarf (Arsenal is a London-based football team) in order to increase their sense of affiliation with the 'victim' (who was an Arsenal supporter) and were warned that the scenario contained strong language. Participants were instructed to: "imagine you are an avid Arsenal supporter; please respond to the situation and characters you meet as you would in everyday life". The experimental scenario lasted for 140 seconds, during which the participants' verbal and behavioural responses were video-recorded. Autonomic arousal and motor activity were indexed throughout the scenario.

In the third stage participants recompleted the PANAS and took part in a semi-structured interview (Appendix 5) about their experiences of the experimental scenario. Lastly, participants engaged in a progressive muscular relaxation exercise (Appendix 5) and listened to a vignette that described the scenario ending amicably between the characters (Appendix 5). This final phase constituted part of the risk management protocol and was specifically intended to reduce any residual arousal or negative feelings participants may have experienced. Participants were debriefed and paid £18.00 (plus travel expenses for the clinical group) for their time. Following participation, the researcher contacted a member of the participants' clinical teams in order to obtain further demographic and clinical information.

2.5 Virtual environment

2.5.1 Apparatus

The VR environment was displayed in an immersive projection system, in which participants were presented with high-resolution images, projected in real-time onto three back-projected wall screens (3 m x 2.2 m) and a floor screen (3 m x 3 m). A stereo presentation of the virtual world was delivered via Lightweight CrystalEyes shutter-glasses worn by participants. These glasses presented separate images to the left and right eyes, producing the illusion of 3D objects both within and beyond the walls of the laboratory. An inertial/ultrasonic head-tracking device was contained within the glasses, which enabled images to be presented with reference to the participants' viewpoint and orientation. This supported almost natural sensorimotor contingencies for visual perception meaning that as participants moved around, the environment displayed perspective-correct information. Spatialised audio was delivered via eight corner speakers which allowed sounds to be appropriately "located", as if emanating from particular objects in the environment.

2.5.2 Environment

The VR environment (Rovira et al., 2009) depicted a virtual pub in which an aggressive interaction escalated between two computer-generated avatars (see Figure 2). Background visual (e.g. pint glasses, bar stools) and auditory (television) stimuli were designed to reflect a typical pub environment. The scenario began with an avatar (V) approaching the participant (X) and establishing a friendly conversation with him about football. Avatar V wore an Arsenal football shirt and principally discussed the Arsenal team with the participant, who had been previously instructed to imagine that they were an Arsenal supporter. In this initial conversation

V talked to participants in a seemingly ad-lib way as illustrated in the following extract taken from Rovira et al's (2009) exploratory study:

V: You alright mate?
X: Oh hello, yes.
V: Where you from?
X: Uh, Kent originally.
V: You're Arsenal yeah?
X: Yea, Yeah sure.
V: Get you!
X (laughing)
V: What did you think of the team last year?
X: Well they got better as the season progressed.
V: Totally agree with ya.

Following this brief conversation, a second male avatar (M) who had been sitting alone at the bar promptly stood up and moved towards V. The conversation between M and V commenced as follows:

M: Hey, you got a problem?
V: Sorry?
M: I said, have you got a problem?
V: No mate.
M: But I saw you looking at me.
V: I didn't look at you.
M: bBut I saw ya… I saw ya staring.
V: No, I wasn't.. I wasn't staring even.
M: Something to get off your chest?
V: No
M: You sure about that?
V: There's nothing wrong mate, there's no trouble, I'm just trying to enjoy a quiet pint.
M: Yeah, that's was I was doing, enjoying a quiet pint
V: Get back to your table and enjoy your quiet pint
M: Don't fucking tell me to go back to my table.

As the scenario progressed, M's behaviour became increasingly threatening both verbally (e.g. shouting and swearing), and non-verbally (e.g. waving arms, invading V's personal space):

M: Why don't you fuck off now? Why don't you leave now?
V: But.. but…
M: Take your pint and fuck off!
V: Why?
M: Because you offend me, that shirt offends me. I told ya, take it off and you can stay here.

V: I'm not gonna take it off.
M: Take it off and you may have a drink.
V: I'm not going…
M: TAKE THE FUCKING SHIRT OFF.
V: Leave it alone… What's going on here?
M: DON'T FUCKING TELL ME TO LEAVE IT OUT! (starts pushing V)
V: What have I done?
M: YOU FUCKING TELL ME TO LEAVE IT OUT? (keeps pushing V)

Throughout the interaction M's body size, gestures and tone indicated a threatening and aggressive demeanour, whereas V appeared submissive and deliberately avoidant of trouble. However V responded during the interaction, M escalated the argument to a more dangerous level. This culminated in M violently pushing V against the wall at which point the scenario ends. Participants were able to respond to M and V both verbally and non-verbally (e.g. reaching out as if to touch an avatar) throughout the altercation but were unable to influence the eventual outcome of the scenario. Participants were unaware that their interventions would be ineffective.

Figure 2: A volunteer with two avatars in the VR pub scenario

2.5.3 Avatars

The avatars were designed to be as realistic as possible and were therefore life-sized (see Figure 2) and displayed in 3D stereo. Two actors provided the avatars' voices and their movements were captured through the use of a Vicon motion picture system. Since participants wore head-tracking equipment, the characters were able to look them in the eye. Avatars were controlled by a hidden operator who selected their utterances and timings from an array of pre-recorded phrases. Typically the conversation followed a predefined pattern, but the system was designed to be flexible enough for the operator to trigger a set of general responses (e.g. "Totally agree with you") when participants said something unexpected during their initial interaction with V.

2.6 Measures

2.6.1 Pre-virtual reality assessment

Basic demographics: Following informed consent, participants provided their age, ethnicity, first language, number of years in education, highest educational qualification and current affiliation with any football team.

The Positive and Negative Affect Schedule (PANAS; Watson, Clark & Tellegen, 1998): The PANAS was administered to measure changes in participants' affective states before and after VR. Consequently participants were instructed to complete it with reference to the present moment. The PANAS is a 20-item measure yielding subscale scores for positive and negative affect. Positive and negative items are presented as single words such as 'excited' and 'hostile', to indicate the extent to which the respondent has felt that emotion within the specified time-frame. Items are rated on a 5-point scale ranging from 1 (very slightly or not at all) to 5 (extremely) with a total score in the range of 20-100. The PANAS possesses robust

reliability and validity for adults, and large-scale normative data is available to assist interpretation (Crawford & Henry, 2004).

2.6.2 Post virtual reality assessment

The PANAS: The PANAS was re-administered following VR in order to measure changes in participants' affective states.

Semi-structured interview: A semi-structured interview schedule (Appendix 5) was developed based on the results of a pilot study in which the researcher conducted an unstructured interview with three volunteers following VR. The interview schedule consisted of 10 questions with additional prompt items, which took approximately 15 minutes to administer. The interview contained two Likert-scale (10-point) items that were designed to gauge the realism of the scenario and the authenticity of participants' behavioural responses. The remaining questions were fixed-choice and open-ended items which focused on participants' thoughts and feelings about the VR experimental scenario. In particular the questions were designed to elicit participants' attributions about the unfolding events, their perceptions of – and feelings towards – the aggressor and victim, and the thoughts that underpinned their emotional responses.

Clinical information: After clinical participants had completed all research tasks, the researcher contacted a member of their clinical team to obtain the following information:

(a) The number and type (e.g. violent, sexual or acquisitive) of previous criminal convictions. This information was gathered by the clinician from past structured risk assessments (e.g. the HCR-20; Webster, Douglas, Eaves & Hart, 1997) and/or Police National Computer records.

(b) Index offence (e.g. the offence that precipitated current contact with mental health services).

(c) Approximate number of years: (i) in contact with mental health services and/or (ii) in receipt of mental health treatment in prison.

(d) Comorbid (i) Axis-I and (ii) Axis-II diagnoses; these had been previously diagnosed according to ICD-10 (WHO, 1992) or DSM-IV (APA, 1994) criteria.

2.7 Data scoring and coding

2.7.1 PANAS data

PANAS positive and negative scales were scored for participants' baseline and post-VR responses. Change scores were calculated for both scales by subtracting post-VR scores from baseline scores. Positive change values reflected increased scores between baseline and VR; negative values reflected reduced scores. Representing PANAS scores both in pre-post format and as change scores made the data suitable for both repeated-measures and correlational analysis.

2.7.2 Video data

Measures of antagonistic/aggressive and conciliatory/prosocial responses were based on ratings of participants' verbal and non-verbal interactions with the two avatars during the experimental scenario. An initial count of verbal and non-verbal behaviours was conducted for each participant using a specifically designed coding chart (Appendix 6), broadly based on the 'intervention criteria' described by Rovira et al. (2009). These scores were then aggregated to produce overall frequencies for (i) aggressive/antagonistic behaviour, and (ii) conciliatory/prosocial behaviours. Ratings were performed by the principal researcher. Fifty percent of videos were randomly selected and independently rated by a blinded member of the research team. Inter-rater reliability was found to be good for conciliatory

interventions: r = 0.89, n = 17, *p*<.001 and antagonistic interventions: r = 0.89, n = 17, *p*<.001.

For scoring purposes, a verbal intervention was defined as any speech act directed towards either avatar during the confrontation. Antagonistic verbal interventions included swearing, using a hostile tone or volume, and provocative comments (e.g. insults) that were likely to exacerbate aggression in real-life contexts. Antagonistic non-verbal behaviours included overt acts of violence (e.g. pushing, punching), aggressive posturing (i.e. squaring-up to an avatar) and provocative gestures (e.g. waving the Arsenal scarf in the aggressor's direction). Conciliatory interventions were defined as behaviours that appeared to function to de-escalate the situation. Verbal conciliatory interventions included pacifying comments (e.g. "Calm down mate") delivered in a calm tone or volume, which were likely to pacify aggression in real life. Conciliatory non-verbal behaviours were those such as standing between the victim and aggressor and making calm gestures (e.g. bringing hands down to signal reductions in arousal).

2.7.3 Interview data

Participants' responses to the two Likert-scale items about authenticity and realism were scored as continuous variables. Since the PANAS was administered as a pre-post measure of affect, it was decided post-hoc that only the four questions concerning cognitions/appraisals from the interview would be analysed and statistically examined. The first of these was a fixed-choice 'victim likability' item (with three response options), which was scored as a categorical variable:

1. *What did you think of the first guy (the victim) when he came in and started chatting to you: did you (a) like, (b) dislike, (c) feel neutral towards him?*

The remaining three 'cognitions' questions were open-ended:

2. *What was going through your mind when the second character (the aggressor) entered and began speaking to the first (the victim)?*

3. *What were the thoughts that were making you feel ……? (repeat the emotion that the participant identified as their dominant emotion in the previous question)*

4. *What was your main motivation for wanting to intervene during the confrontation (e.g. to intimidate aggressor/pacify/protect victim etc.)?*

Participants' responses to these three questions were content analysed. This approach assumes that the many words in text can be reduced into much fewer content categories, whilst maintaining their original themes, issues and ideas (Weber, 1990). Emergent categories can then be analysed using standard statistical methods. Accordingly, participants' verbatim responses were coded into an overarching emergent theme of 'cognitive salience', constituting three broad categories: (a) victim-salient, (b) aggressor-salient and (c) self-salient. Victim-salient responses reflected thoughts of concern for the victim's wellbeing/safety, wanting to help/rescue him, expressing empathy (e.g. fearful on his behalf) or guilt for not helping. The latter two categories reflected anger-oriented cognitive appraisals and interpretations. Specifically, aggressor-salient responses denoted hostile cognitions towards the aggressor, including derogatory descriptions of him, expressing a desire to physically or verbally attack him, or wanting to teach him a lesson. Self-salient cognitions were those that concerned the participant such as thoughts of being personally affronted, humiliated, disrespected and/or undermined by the confrontation, or frustrated at being unable to effect change in the outcome of the scenario.

For each participant, a total score was computed for each type of cognition (victim-salient, aggressor-salient and self-salient). This was calculated by aggregating participants' responses to each of the three interview questions. Participants could therefore achieve a score for each type of cognition in the range of 0-3. This nominal data was entered into SPSS for non-parametric analysis.

2.8 Statistical analysis

To address the study's hypotheses, the data was analysed in the following steps:

1. *Affect*: Mixed-model ANOVAs were computed to assess whether changes in participants' positive and negative affective states from baseline to post-VR differed according to diagnostic status.

2. *Cognitions*: Mann-Witney U tests were computed to assess whether the frequency of each type of cognition (pooled across the three interview questions) differed between the two groups. These were conducted separately for each type of cognition (victim-salient, aggressor-salient and self-salient). For the 'victim likeability' item, a chi-square test was computed with diagnostic status.

3. *Behaviour*: T-tests were computed to assess whether the frequency of participants' behavioural responses (antagonistic and conciliatory) differed according to diagnostic status.

4. *Relationship between dependent outcomes*: Two sets of correlational analyses were conducted in order to examine the relationship between behaviour and emotional reactivity. First, Pearson's correlations were computed between PANAS (positive and negative) change scores and the frequency of behavioural interventions (antagonistic and conciliatory). Second, Spearman's Rho correlations were computed between the frequencies of behavioural interventions and the three types of nominal 'cognitions'.

3. RESULTS

3.1 Data preparation

Prior to analysis, the distributions of all variables were evaluated against parametric test assumptions. Variables were examined separately for the 15 ASPD participants and 20 healthy volunteers. Bar charts of the three categorical 'cognitions' variables confirmed that frequency splits fell within acceptable limits for both ASPD and healthy volunteer groups. Frequency splits were evaluated against Rummel's (1970) criteria that only 90-10 splits (or above) are problematic because the correlation coefficients between these and other variables are truncated.

Visual depictions of all continuous outcome variables revealed one high-scoring ASPD outlier on the PANAS negative scale at baseline, two low-scoring ASPD outliers for the PANAS positive scale at baseline and one high-scoring ASPD outlier for antagonistic behaviours. One healthy volunteer high-scoring outlier was observed for the PANAS positive scale post-VR, and two others for antagonistic behavioural interventions. Outliers across variables appeared to reflect the general direction of scores obtained by other cases in their respective groups. All cases were therefore considered legitimate for retention in the analyses but steps were taken to reduce their impact. In accordance with Tabachnick & Fidell's (2001) recommendations, outliers were assigned a raw score on the offending variable(s) that was one unit (raw score) larger (or smaller) than the next most extreme score in the distribution.

Normality of variables was assessed via visual inspection of histograms and by checking skewness and kurtosis values. For the healthy volunteer group, antagonistic interventions were initially observed to be non-normally distributed, owing to mild skewness ($z = 3.54$). However once the outlier for this variable had

been assigned a less extreme value, skewness fell within the margins of acceptable limits (z = 3.02). All other variables were normally distributed.

3.2 Demographic information

The demographic characteristics of the two study groups are presented in Table 1. Clinical participants had a mean age of 44 (*SD* = 10.10) and healthy volunteers of 28 (*SD* = 7.11). Most clinical participants left school without any educational qualifications (87%), whereas all healthy volunteers had obtained higher educational qualifications. Of the clinical participants 60% were White British, compared with 40% of healthy volunteers.

Table 2 displays clinical and criminogenic characteristics of the ASPD group. Consistent with comorbidity rates in community samples (e.g. Robins & Price, 1991), the majority (74%) had at least one comorbid Axis I diagnosis, most commonly substance misuse (60%). More than half (53%) had a comorbid Axis II diagnosis, the majority being borderline personality disorder. Clinical participants had serious and extensive offending histories, with an average of 23 previous convictions and the majority of index offences being violent (53%) or sexual (27%) in nature.

Table 1: Participant demographics

	ASPD (n =15)	Healthy volunteers (n = 20)	Statistic	*p* value
Age in years: *M* (*SD*)	43.7 (10.10) range: 24-62	27.8 (7.11) range: 20-47	Z = -3.86	>.001*
Education years: *M* (*SD*)	9.3 (1.60) range: 6-13	19.1 (2.97) range: 14-24	Z = -5.05	>.001*
Highest qualification: *N* (%)				
None	13 (86.7%)	-		
GCSEs	1(6.7%)	-		
A-levels	1 (6.7%)	-		
Higher education	-	20 (100.0%)		
First language: *N* (%)			χ^2 =3.29	0.09
English	15 (100%)	16 (80%)		
Non-English	-	4 (20%)		
Ethnicity: *N* (%)				
White British	9 (60.0%)	8 (40.0%)		
Asian	-	5 (25.0%)		
Black	1 (6.7%)	2 (10.0%)		
Mixed	3 (20.0%)	1 (5.0%)		
Chinese or other	-	1 (5.0%)		
White other	2 (13.3%)	3 (15.0%)		
Football team: *N* (%)			χ^2 = 3.07	0.21
Arsenal	1 (6.7%)	6 (30.0%)		
Other	5 (33.3%)	6 (30.0%)		
None	9 (60.0%)	8 (40.0%)		

Table 2: Clinical and criminogenic characteristics of the ASPD group

Characteristic	Descriptive data
Number of Axis I disorders: *M* (*SD*)	1.10 (0.88) range: 0-3
Type of Axis I disorder: *N* (%)	
Schizophrenia	2 (13.3%)
Major Depression	2 (13.3%)
Anxiety disorder	3 (20.0%)
Substance misuse	9 (60.0%)
Number of Axis II disorders: *M* (*SD*)	0.6 (0.63)
Type of Axis II disorder: *N* (%)	range: 0-2
Borderline	7 (46.7%)
Schizotypal	1 (6.7%)
Paranoid	1 (6.7%)
Index offence: *N* (%)	
Violent	8 (53.3%)
Sexual	4 (26.7%)
Acquisitive	3 (20.0%)
Number previous convictions: *M* (*SD*)	20.3 (24.0) range: 1-95
Treatment in mental health services/prison (years): *M* (*SD*)	11.7 (7.72) range: 3-29

3.3. Baseline equivalence of groups

Between-group equivalence tests (see Table 1) revealed that the ASPD group was older and less educated than healthy volunteers. This most likely reflected the fact that ASPD participants had invariably spent a number of years in secure facilities (and were therefore older), and lower educational level is an inherent characteristic of ASPD (Robins, Tipp & Pzybeck, 1991). Further analysis indicated that age and education correlated with several dependent outcomes for the two groups combined (see Table 3). Although it was considered unlikely that these demographic characteristics were responsible for between-group differences in outcomes (especially given that research suggests that emotion regulation increases with age; Urry & Gross, 2010), further education- and age-effects

analyses were computed for healthy volunteers as a precautionary safeguard (see Table 3). This confirmed that neither variable was associated with any of the dependent outcomes. Therefore further analysis did not adjust for baseline differences.

Table 3: Correlations between age, education and outcome variables

		Age	Education
Combined groups			
PANAS negative scale change scores	Pearson	0.43*	-0.52**
PANAS positive scale change scores	Pearson	0.14	0.01
Total antagonistic behaviours	Pearson	0.46**	-0.36*
Total conciliatory behaviours	Pearson	-0.29	-0.29
Total victim-salient cognitions	Spearman	-0.32	-0.50**
Total aggressor-salient cognitions	Spearman	0.19	-0.31
Total self-salient cognitions	Spearman	0.39*	-0.44**
Healthy volunteers only			
PANAS negative scale change scores	Pearson	0.22	-0.19
PANAS positive scale change scores	Pearson	0.42	-0.42
Total antagonistic behaviours	Pearson	0.05	-0.04
Total conciliatory behaviours	Pearson	0.22	-0.18
Total victim-salient cognitions	Spearman	-0.03	0.03
Total aggressor-salient cognitions	Spearman	0.03	-0.03
Total self-salient cognitions	Spearman	-	-

Note *p< .05 level
Note **p< .01
Note - correlation not applicable: frequency of responses = zero

3.3 VR scenario ecological validity

Interview question: *How realistic were the unfolding events in the experience that you have just had (10-point Likert scale)?* Descriptive data from the two groups combined indicated that the scenario was perceived to be sufficiently realistic (*M* = 7.50, *SD* = 1.74). ASPD participants found the scenario more realistic (*M* = 8.33, *SD* = 1.45) than healthy volunteers (*M* = 6.90, *SD* = 1.71): $t(33) = 2.61$; $p = 0.013$.

Interview question: *To what extent was your reaction authentic/the same as you would do in real life (10-point Likert scale)?* Overall, participants reported that their

reactions to the VR scenario were quite authentic (*M* = 7.49, *SD* = 1.82). There was no difference between the two groups, with ASPD participants reporting similar levels of authenticity (*M* = 7.73 *SD* = 1.49) to healthy volunteers (*M* = 7.30, *SD* = 2.10): *t*(33) = 0.69; *p* = .49.

3.4 Group comparisons

3.4.1 The affect hypothesis was that between baseline and post-VR, ASPD participants would show a greater increase on the negative scale and a greater decrease on the positive scale of the PANAS than healthy volunteers.

PANAS negative scale: PANAS negative scores violated Mauchly's Test of Sphericity (χ^2 = 0.00; *p*<.01); therefore ANOVA results were interpreted via the Greenhouse-Geisser test. Descriptive PANAS data is presented in Table 4. There was a main effect of 'point of measurement' (baseline – post-VR), indicating that for the two groups combined, scores increased from baseline (*M* = 15.00) to post-VR (*M* = 19.34): F(1, 33) = 25.50; p > 0.001 (*d* = 1.75). There was no main effect of diagnostic group; ASPD participants did not differ in their PANAS negative scores averaged across the two time points (*M* = 17.93) compared with healthy volunteers (*M* = 16.60): *F*(1,32) = 0.75; *p* = .39. In line with the hypothesis, there was a significant interaction (see Figure 3) between 'point of measurement' and 'diagnostic status', confirming that ASPD participants showed a greater increase on the PANAS negative scale between baseline and post-VR (mean increase = 8.40) than healthy volunteers (mean increase = 1.30): *F*(1,33) = 13.66; *p* < 0.001 (*d* = 1.28)

Table 4: PANAS scores at baseline and post-VR

Diagnostic status		Point of measurement: Baseline *M* (*SD*)	Point of measurement: Post-VR *M* (*SD*)
ASPD (*n* = 15)	PANAS negative	13.73 (3.37)	22.13 (6.25)
	PANAS positive	35.93 (3.97)	33.20 (8.04)
Healthy volunteers (*n* = 20)	PANAS negative	15.95 (5.76)	17.25 (5.27)
	PANAS positive	29.40 (6.68)	29.20 (7.42)

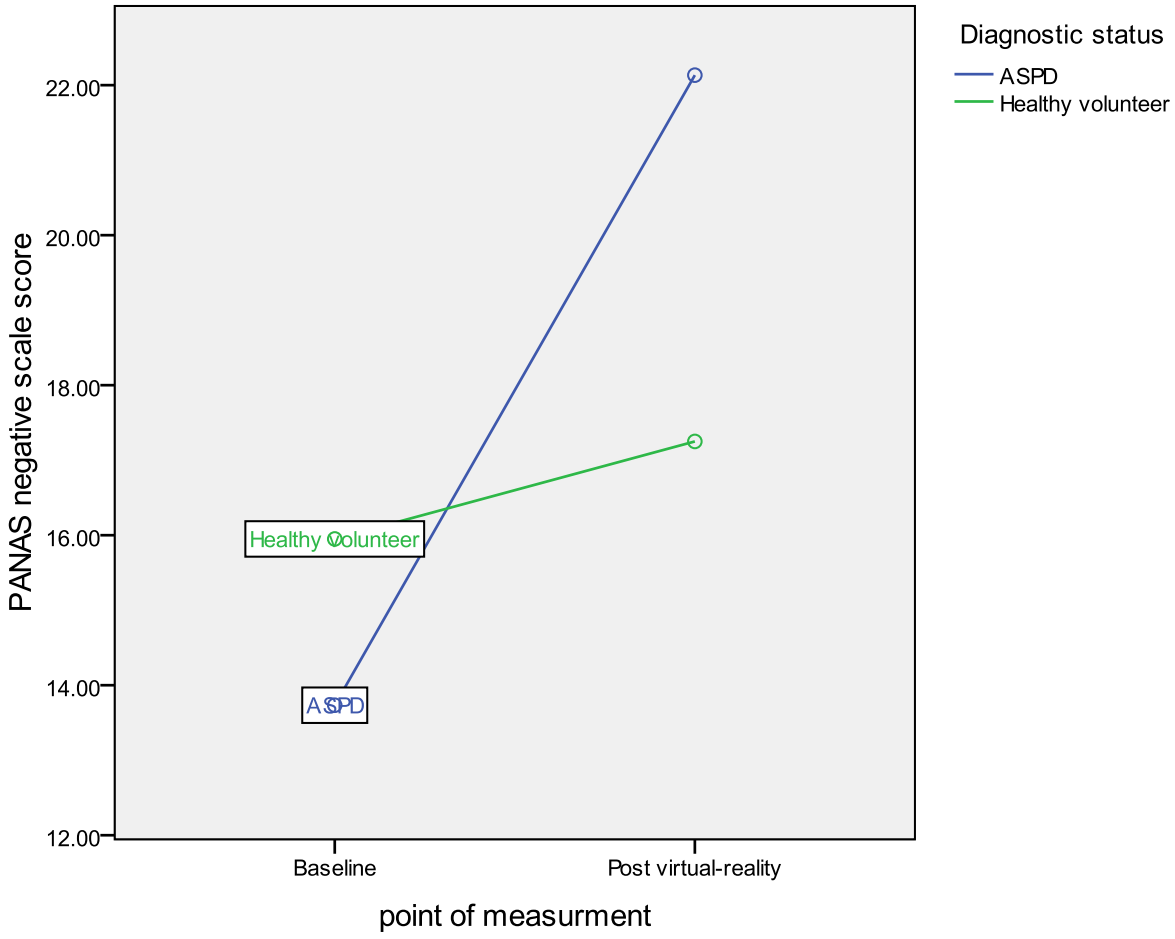


Figure 3: Diagnostic status by point of measurement for PANAS negative scale

PANAS positive scale: PANAS positive scale scores (see Table 4) violated Mauchly's Test of Sphericity (χ^2 = 0.00; p<.01): therefore ANOVA results were interpreted via the Greenhouse-Geisser test. There was no main effect of 'point of

measurement' (baseline – post-VR), indicating that for the two groups combined, there was no difference between PANAS positive scores at baseline (*M* = 32.20) compared with post-VR (*M* = 30.91): *F*(1, 33) = 1.91; *p* = .18. There was a main effect of 'diagnostic status' (ASPD – healthy volunteers), indicating that ASPD participants yielded higher scores averaged across the two time points (*M* = 34.57) compared with healthy volunteers (*M* = 29.30): *F*(1,33) = 6.58; *p* = .02 (*d* = 0.81). Contrary to the hypothesis, there was no interaction between 'diagnostic status' and 'point of measurement', indicating that the groups did not differ from one another in their PANAS positive scale change scores: *F*(1, 33) = 1.43; *p* = .24.

3.4.2 The cognitions hypothesis was that participants with ASPD would report more anger-oriented and less victim-empathic cognitions about the VR scenario than healthy volunteers.

Across the three combined 'cognitions' interview questions, ASPD participants reported more anger-oriented cognitions than healthy volunteers. Specifically, Table 5 shows that the ASPD group reported more 'aggressor-salient' cognitions (*Mdn* = 1) than healthy volunteers (*Mdn* = 0): U = 97.0, z = -1.97, *p* = .02 (*r* = 0.33). The ASPD group also reported more 'self-salient' cognitions (*Mdn* = 1) than healthy volunteers (*Mdn* = 0): U = 70.0, z = -3.64, *p* < .001 (*r* = 0.62). Conversely, ASPD participants endorsed fewer 'victim-salient' cognitions (*Mdn* = 1) than healthy volunteers (*Mdn* = 3): U = 52.5, z = -3.48, *p* < .001 (*r* = 0.59). These findings were consistent with the study's hypothesis.

Table 5: Descriptive and significance data for cognitions and behaviour outcomes

Dependent outcome	ASPD N = 15	Healthy volunteers N =20	Statistic	Sig	E.S[1]
Cognitions	*Mdn* (Range)	*Mdn* (range)	U	*p*	*r*
Victim-salient	1.0 (1-3)	3 (1-3)	52.5	.001**	0.59
Aggressor-salient	1.0 (1-3)	0 (0-1)	97.0	.02*	0.33
Self-salient	1.0 (1-3)	0 (0)	70.0	<.001*	0.62
Behaviour	*M* (*SD*)	*M* (*SD*)	t	*p*	*d*
Conciliatory	2.80 (2.24)	7.05 (4.52)	-3.34	.001**	1.26
Antagonistic	2.40 (3.46)	0.70 (1.22)	1.82	.08	0.73

Note * $p < .05$ level (one-tailed)
Note ** $p < .01$ level (one-tailed)

There was a moderately strong association between 'victim likeability' and diagnostic status: χ^2 (2, n = 35) = 10.21; p = .006, Cramer's V = 0.54. Table 6 shows that ASPD participants were more prone to disliking the victim than were healthy volunteers (40% and 0%, respectively). Conversely healthy volunteers were more likely to report liking the victim than were ASPD participants (65% and 13%, respectively). These findings were in line with the study's hypotheses and consistent with the aforementioned finding that ASPD participants reported fewer victim-salient (empathic) cognitions than healthy volunteers.

Table 6: Descriptive data for 'victim likeability'

	ASPD (n = 15)	Healthy volunteers (n = 20)	Total (N = 35)
Cognition type	*n* (%)	*n* (%)	*N* (%)
Liked victim	2 (13%)	13 (65%)	15 (43%)
Disliked victim	6 (40%)	0 (0%)	6 (17%)
Neutral to victim	7 (46%)	7 (35%)	14 (40%)

[1] For cognitions effect size calculations r = Z / √N. Values were evaluated against Cohen's (1988) criteria where *r*: small ≥ .10, medium ≥ .30, large ≥ .50.

3.4.3 The behaviour hypothesis was that participants with ASPD will display more aggressive/antagonistic behavioural interventions and fewer conciliatory/prosocial interventions during VR than healthy volunteers.

Significant heterogeneity of variance was observed for both antagonistic (Levene $F(1, 33) = 15.85$; $p < .001$) and conciliatory interventions (Levene $F(1, 33) = 7.75$; $p = .009$). Therefore t-tests were interpreted via 'equal variances not assumed' and non-parametric tests (Mann-Whitney U) were computed as an adjunct. Table 5 shows that there was a trend towards ASPD participants conducting more antagonistic/aggressive interventions ($M = 2.40$) than healthy volunteers ($M = 0.70$); however this did not reach significance on parametric: $t(33) = 1.82$; $p = .08$, and non-parametric tests: U = 118.5; $p = .11$ (one-tailed). Participants with ASPD conducted fewer conciliatory/pro-social behavioural interventions ($M = 2.80$) than healthy volunteers ($M = 7.05$): $t(33) = -3.34$; $p = .001$ ($d = 1.26$), U = 67.5; $p = .006$ (one-tailed).

3.2.4 The hypothesised relationship between outcomes was that antagonistic behaviour would positively correlate with negative affect and anger-oriented cognitions. Conversely, conciliatory behaviour was expected to positively correlate with positive affect and victim-empathic cognitions.

Behaviour and affect: In accordance with the hypotheses, increased negative affect positively correlated with more antagonistic behaviour: $r = 0.348$, $N = 35$; $p = .02$, and increased positive affect positively correlated with conciliatory behaviour: $r = .466$, $N = 35$; $p = .002$ (see Table 7).

Behaviour and cognitions: In accordance with the hypotheses, there was a positive correlation between empathy-oriented (victim-salient) cognitions and conciliatory behaviour: r = 0.284, N = 35; p = 0.029. Table 7 shows that whilst trends concerning the relationship between anger-oriented cognitions (aggressor-salient

and self-salient) and behaviour were in the direction of the hypothesis, they did not reach statistical significance.

Table 7: Correlations between behaviour and other outcomes

		Conciliatory behaviour	Antagonistic behaviour
Dependent outcome	Correlation		
PANAS negative scale change score	Pearson's	-	0.348*****
PANAS positive scale change score	Pearson's	0.466******	-
Empathy-oriented cognitions			
Victim-salient	Spearman's Rho	0.284*****	-
Anger-oriented cognitions			
Aggressor-salient	Spearman's Rho	-	0.268
Self-salient	Spearman's Rho	-	0.274

Note ***** p<.05 (one-tailed)
Note **p<.01 (one-tailed)
Note - correlation not computed

4. DISCUSSION

4.1. Main findings

The current study investigated the influence of VR provocation on the emotional and behavioural reactivity of participants diagnosed with ASPD compared with healthy volunteers. The results suggested that the VR experimental scenario was an ecologically valid and effective means of inducing negative emotionality: participants reported finding the scenario – and their responses to it – quite authentic, and showed an increase in negative affect from baseline to post-VR.

There are three main findings with regard to the overarching hypothesis that provocation would result in greater negative emotional and behavioural reactivity in ASPD participants than in healthy volunteers. The ASPD group showed (1) greater increases in negative affect, (2) more anger-oriented and fewer empathy-oriented cognitions, and (3) less conciliatory/prosocial behaviour than the healthy volunteers. ASPD participants also displayed a trend towards increased aggressive behaviour. Modest support was provided for the subsidiary hypothesis: prosocial behaviour was correlated with positive affect and victim-empathic cognitions but the relationship between aggression and negative emotional reactivity was less compelling, with several non-significant trends. Contrary to predictions, there was no evidence that the VR scenario impacted on participants' positive affective states, either within or between groups.

4.1.1 Negative emotional reactivity

Between-group differences in negative affective and cognitive reactivity provide tentative support for the notion that ASPD may entail difficulties regulating emotions under conditions of perceived threat. This is in line with previous research indicating that impulsive traits of ASPD are associated with negative emotional

reactivity (Hicks & Patrick, 2006) and with deficits in brain regions that govern inhibition (Morgan & Lilienfeld, 2000).

Comparative increases in anger-oriented cognitions mirror Lobbestael et al's (2009) observation of cognitive hyper-responsivity in ASPD compared with control participants. However, the present finding that ASPD participants displayed greater increases in negative affect was not observed by Lobbestael et al. (2009). This discrepancy could be accounted for by incomparable affective outcome measures; the present study indexed a broad dimension of negative affect (using the PANAS) whilst Lobbestael et al. (2009) measured only anger. It is possible that individuals with ASPD experience a range of negative affective changes (e.g. shame, humiliation) in response to provocation, which narrower measures of anger fail to capture.

A further explanation for the above inconsistency lies in the notion that the strength and imminence of an anger stimulus determines whether an antisocial individual displays an 'orienting' or 'fight' response in their defensive shifting (Fanselow & Lester, 1988). It is speculated that because the present research utilised a more imminent (and ecologically valid) anger-induction method than Lobbestael et al. (2009), ASPD participants experienced affective shifts in accordance with a defensive response. Taken together these studies raise the possibility that individuals with ASPD display more predatory-like orienting strategies under conditions of remote threat, but atypical increases in negative emotional reactivity when imminent threat is perceived. Future studies should aim to test this hypothesis via remote and imminent VR provocation, incorporating physiological measurement as an objective dimension of defensive reactivity.

4.1.2 Emotional reactivity and conciliatory/prosocial behaviour

Between-group differences in conciliatory behaviour – and the observed relationship between these behaviours and indices of positive emotionality (cognitive and affective) – tentatively suggest that patterns of emotional reactivity influenced expressions of prosocial behaviour. These findings are in line with developments in moral psychology implicating emotions and empathy as essential ingredients of prosocial behaviour (Gilligan, 1993). However, following provocation no differences were observed between the groups in positive affect, suggesting that reductions in positive emotionality alone are unlikely to account for less conciliatory behaviour in the ASPD group.

It is tentatively suggested that the paucity of prosocial interventions reflected a combination of reduced positive emotionality (i.e. less victim-empathic cognitions) and increased negative emotionality (i.e. more self-salient and aggressor-salient cognitions) arising from pre-existing deficits in the ASPD group. This is consistent with research showing that ASPD entails subtle impairments in empathy and perspective-taking (Dolan & Fullam, 2004) and that reduced prosocial behaviour is associated with a lack of appreciation for the emotional states of others (Blackhart, Nelson, Knowles, & Baumeister, 2009), poor emotion regulation (Twenge, Baumeister, DeWall, Ciarocco, & Bartels, 2007) and increased focus on the self (DeWall & Baumeister, 2006).

Conceptualised within a mentalization model of ASPD (Bateman & Fonagy, 2008) mentalization deficits are exacerbated when the attachment or affiliative system is activated, meaning that individuals with ASPD experience their own actions as being without consequence and become unable to extricate another's mental state. Thus comparatively less conciliatory behaviour from ASPD participants could be formulated as a temporary inability to experience the victim's perspective and a reduced sense of their own ability to effect change during a period of

interpersonally provoked emotional hyper-arousal. However, these inferences are speculative and further research is required to directly examine the relationship between emotional reactivity, mentalizing and the commission of prosocial behaviour in ASPD.

Alternatively, it is possible that individuals with APSD would fail to display prosocial behaviour even during periods of emotional stability, when no direct threat to self is experienced. Research shows that adolescents with externalising problems express less affiliative (relationship-building) social goals than healthy volunteers (Lochman, Wayland & White, 1993) and young offenders display less mature moral judgments than their peers (Larden, Melin, Holst & Langstrom, 2006). It is plausible that these types of baseline deficits would result in less prosocial behaviour in ASPD, regardless of levels of emotional arousal. Future research could include a baseline activation task (e.g. a non-anger-inducing VR scenario designed to elicit prosocial behaviour and moral decision-making) to investigate the extent to which prosocial behaviour deficits in ASPD originate from pre-primed social processing deficits versus emotional activation.

An important factor to consider when interpreting prosocial behavioural outcomes is that the clinical participants had been in receipt of mental health and/or prison treatment for an average of 12 years. Treatment for violent offenders typically comprises cognitive-behaviourally informed anger management (e.g. Violent Offender Treatment Protocol; Braham, Jones & Hollin, 2008), which seeks to promote contingency management and behavioural inhibition skills. Thus a failure to intervene could be seen as an adaptive learnt strategy for ‘staying out of trouble’ (i.e. removing oneself from involvement in conflict). This – and other methodological factors (see below) – complicates interpretation.

4.1.3 Emotional reactivity and antagonistic/aggressive behaviour

Trends regarding antagonistic/aggressive behaviour provide tentative support for the notion that as with other cluster B disorders, persons with ASPD may have particular problems inhibiting aggressive impulses during emotionally laden situations (Verona, Sprague & Sadeh 2012). Findings are consistent with research demonstrating that emotional contexts can impede the ability to inhibit aggression (Verona & Sedeh & Curtin, 2009), sustained negative affect biases high stress-reactive individuals towards more intense and sustained acts of aggression (Verona et al., 2002) and individuals with ASPD show high levels of reactivity and prefrontal cortex deficits when exposed to negative emotional contexts (Sprague & Verona, 2010).

In a similar vein to prosocial outcomes, these findings could be formulated within a mentalization model of ASPD (Bateman & Fonagy, 2008), which suggests that the common pathway to violence is a momentary inhibition in the capacity for mentalization, triggered by a sense of threat to self. Crucially, this model takes account of a diminished sense of self-worth, which previous explanatory models of ASPD argue underpins violence, by way of shame that can be dealt with only through the violent humiliation of others (e.g. Gilligan, 1996). Qualitatively, this fits with the greater frequency of self-salient and aggressor-salient cognitions and a trend towards increased aggression observed in the current ASPD sample. However, in light of the exploratory nature of the study and the fact that mentalization was not directly indexed these inferences are speculative. Furthermore, the absence of statistical significance means that particular caution should be applied not to over-interpret aggressive behavioural outcomes.

There are a number of potential explanations for between-group differences in aggressive behaviour failing to reach significance. First, as discussed in relation to conciliatory behaviour, clinical participants may have honed skills in treatment,

enabling them to inhibit aggressive impulses in spite of increased negative emotionality. Second, although participants reported that the scenario and their responses to it were quite authentic, it is possible that 'in vitro' provocation lacked the necessary properties to provoke a full-blown fight response. It is unfortunate that the physiological data collected proved uninterpretable as this would have facilitated a more proficient evaluation of the scenario's anger-arousal potential. Finally, it is of course possible that individuals with ASPD are no more likely to respond impulsively to provocation than non-aggression-prone persons, and aggression in ASPD may be more akin to the instrumental type observed in psychopathy. These competing explanations, coupled with methodological and design limitations detailed below, complicate interpretation of the findings.

4.2 Limitations

Several limitations warrant discussion. First with regard to generalisability, the degree to which the clinical group was representative of the ASPD population is questionable owing to the aforementioned length of time spent in treatment services, the fact that only patients with community leave (indicative of risk reduction) were eligible to take part, and the relatively high refusal rate. Consequently it is possible that the achieved sample posed less risk of aggression than individuals with ASPD in the general population. Alternatively, because recruitment primarily took place in specialist forensic personality disorder services, individuals with particularly profound personality disorders and forensic needs may have been recruited. The fact that participants had an average of 20 previous convictions and index offences of a primarily violent or sexual nature supports this possibility. Future studies are required with untreated ASPD individuals (i.e. upon entry to services) in order to better understand patterns of emotional reactivity inherent in this population, and establish whether these are ameliorated with treatment. External validity may have

also been compromised by the fact that the healthy volunteers largely comprised students with higher educational qualifications.

A second limitation in relation to sampling is that the majority of clinical participants had comorbid Axis I and Axis II diagnoses, most commonly substance misuse and borderline personality disorder. Given that both of these diagnoses are associated with difficulties in emotion regulation, it is not possible to disentangle how much of the variability in emotional and behavioural reactivity was attributable to ASPD symptomatology. Future research should endeavour to address this by screening out comorbid psychiatric conditions and/or including control groups. Also problematic is that rates of psychopathy were not assessed and it is likely that some clinical participants – particularly those recruited from specialist forensic personality disorder services – met criteria for psychopathy. Given that psychopathic offenders show chronic under-arousal during provocation (e.g. Lobbestael et al., 2011), the responses of more callous/unemotional participants could have reduced between-group differences in emotional and behavioural reactivity. Controlling for psychopathy represents an important priority in the event of replication.

A noteworthy potential confounding factor is that the ASPD group found the experimental scenario more realistic than the healthy volunteers. Qualitative reports during interview suggested that this was because many clinical participants had previously been involved in similar conflicts in pubs. It is possible that these prior experiences influenced emotional responsivity (i.e. akin to a trauma response) and future studies should endeavour to design VR scenarios that are equitable in salience for clinical and healthy volunteer groups. Furthermore, as aforementioned this study did not include a baseline activation task, meaning it was not possible to differentiate how much of the variance in emotional arousal resulted from differences in emotional processing versus emotional reactivity. Additionally, the

experimenter was not blinded to group membership and may have therefore unwittingly influenced participants' behaviour during VR and interview responses.

Several limitations should also be acknowledged in relation to measurement. First, behavioural reactivity was assessed during VR provocation, whereas self-reported affect and cognitions were indexed immediately afterwards. This was done in order to capture behaviour 'live' but means that a time effect could have confounded the comparison between behaviour and indices of emotional reactivity. This is one potential explanation for the correlations between anger-oriented cognitions and aggressive behaviour falling short of significance. Second, as aforementioned the PANAS was administered in order to assess broad dimensions of positive and negative affect, but this meant it was unclear which feelings changed following provocation. Third, reliance on self-report tools is problematic in antisocial populations (Lobbestael et al., 2009) and future research should endeavour to obtain physiological markers of emotional reactivity as originally intended. Fourth, a semi-structured interview was developed with the aim of eliciting participants' experiential perspectives but the post-hoc decision to only analyse cognitions ultimately resulted in a somewhat reductionist approach. Finally, multiple statistical tests were conducted thereby increasing the Type I error rate. A Bonferoni corrective procedure was not applied as this was an exploratory study and it was considered important to identify any potential relationships in the data for future researchers to investigate. It is somewhat reassuring that the majority of effect sizes were medium or large, suggesting that significant findings are likely to be replicable.

4.3 Implications for research and practice

This study highlights the need for continued large-scale and methodologically robust research into the emotional reactivity of individuals with ASPD. Future studies should endeavour to address the current limitations by

controlling for comorbidity, administering a baseline VR activation task, developing a scenario that is equally salient for clinical and control groups, ensuring blinding of outcome assessors and indexing physiological markers of arousal. In particular research is required to better understand the effects of remote versus imminent threat, patterns of reactivity in untreated ASPD samples, and the relationship between mentalization, emotional reactivity, prosocial behaviour and aggression. VR represents a promising method for pursuing these objectives but the anger-arousal potential of 'in vitro' provocation requires more precise evaluation.

The results of this study hold several implications for the treatment of persons with ASPD. If, as the findings suggest, interpersonal aggression primarily reflects emotional hyper-responsivity, then treatments that target emotional and behavioural regulation may be indicated, such as dialectical behaviour therapy (Linehan, 1987) and mentalization-based treatment for ASPD (Bateman & Fonagy, 2008). Controlled research into the efficacy of these approaches is an important priority, particularly given that a recent systemic review highlighted "insufficient trial evidence to justify using any form of psychological intervention with adults with ASPD" (Gibbon, Duggan, Stoffers, et al., 2010).

The current findings suggest that in addition to facilitating a better understanding of the processes involved in ASPD, VR may prove useful in the treatment of aggression and violence. At present, cognitive-behavioural therapy is the most common approach to anger-management (Beck & Fernandez, 1998) and there is some – albeit limited – evidence of efficacy for improving outcomes in ASPD (Gibbon et al., 2010). This approach emphasises the development and rehearsal of adaptive skills, traditionally honed through imaginal or role-play exposure. Recent research suggests that VR may be a potent alternative for anger exposure in non-clinical groups (e.g. Miyahira et al., 2010) and the current findings suggest that individuals with ASPD may also benefit from 'in vitro' methods. Involving persons diagnosed with ASPD in the design of VR treatment scenarios may help maximise

ecological validity and promote service-user involvement, which is currently lacking in forensic services (National Service User Involvement Network, 2011).

A further potential use for VR concerns risk assessment in forensic personality disorder services. An ecologically valid scenario, sufficiently potent to provoke authentic responses, may enable a 'risk algorithm' to be generated based on the four markers of affect, cognitions, physiological arousal and behaviour. For example the current finding that 40 percent of ASPD participants disliked the victim avatar and none of the healthy volunteers expressed similar thoughts, suggests that negative attitudes towards a VR victim (coupled with other indicators of reactivity) could indicate a propensity towards antisocial behaviour. However, the small sample size and other methodological drawbacks in the present study, renders these inferences premature. Modified large-scale replications are required to ascertain the utility, feasibility and predictive validity of virtual risk assessments. The latter could be evaluated by comparing the longitudinal accuracy of VR probability risk estimates to those obtained with established paper-based tools (e.g. the HCR-20; Webster et al., 1997).

A final implication for practice concerns attitudes toward ASPD. Individuals with this diagnosis are stigmatised and have historically been excluded from services (NICE CG 77, 2010). Negative attitudes are likely to be exacerbated by the terms ASPD and psychopathy being used interchangeably, despite the diagnoses being clinically distinct. Conceptualising ASPD as a disorder that entails difficulties in regulating emotions may engender more optimistic and less pejorative attitudes in staff working with this population, and provide an alternative discourse to the 'cold-blooded' offender frequently portrayed in the media. Reducing stigma and discrimination in ASPD is an important policy objective (Personality Disorder: no longer a diagnosis of exclusion; DoH, 2003) and reflected in treatment guidelines (NICE CG 77, 2010).

4.4 Summary

In summary this is the first study known to have assessed the impact of provocation on individuals with a well-established diagnosis of ASPD, using a seemingly ecologically valid stimulus. Overall, clinical participants showed greater negative emotional reactivity, less victim empathy and conciliatory behaviour, and a trend toward more aggression. Findings support the notion that ASPD entails difficulties with emotion regulation and behavioural inhibition under conditions of perceived imminent threat. Potential implications include ascertaining the efficacy of emotion-regulation-driven treatment approaches for ASPD, reinforcing the distinction between ASPD and psychopathy, and using VR as a tool to assess and treat risk of aggression/violence in this population. However, these findings are exploratory and large-scale modified replications are required before firm conclusions can be reached about patterns of emotional and behavioural reactivity in ASPD.

REFERENCES

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders, (Revised 4th ed.)*. Washington DC: APA.

Bateman, A., & Fonagy, P. (2008). Comorbid antisocial and borderline personality disorders: mentalization-based treatment. *Journal of Clinical Psychology, 64,* 181-194.

Beck, R., & Fernandez, E. (1998). Cognitive-behavioral treatment of anger: A Meta-analysis. *Cognitive Therapy and Research, 22*, 63-74.

Berkowitz, L. (1990). On the formation and regulation of anger and aggression: A cognitive-neoassociationistic analysis. *American Psychologist, 45,* 494–503.

Berkowitz, L. (1983). Aversively stimulated aggression: Some parallels and differences in research with animals and humans. *American Psychologist, 38,* 1135-1144.

Berkowitz, L. (1994). Is something missing? Some observations prompted by the cognitive-neoassociationist view of anger and emotional aggression. In L. R. Huesmann (Ed.), *Aggressive behavior: Current perspectives* (pp. 35–57). New York: Plenum.

Blackhart, G.C., Nelson, B.C., Knowles, M.L., & Baumeister, R.F. (2009). Rejection elicits emotional reactions but neither causes immediate distress nor lowers self-esteem: A meta-analytic review of 192 studies on social exclusion. *Personality and Social Psychology Review, 13*, 269-309.

Blake, P.Y., Pincus, J.H., & Buckner, C. (1995). Neurologic Abnormalities in Murderers. *Neurology*, *45*, 1641-1647.

Bohil, C.J., Alicea, B., & Biocca, F.A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, 12, 752-762.

Braham, L., Jones, D., & Hollin, C.R. (2008). The Violent Offender Treatment Program (VOTP): Development of a treatment programme for violent patients in a high security psychiatric hospital. *International Journal of Forensic Mental Health*, *7*, 157-172.

Brinkman, W.P., Hattangadi, N., Meziane, Z., & Pul, P. (2011). *Design and evaluation of a virtual environment for the treatment of anger.* Paper presented at the Virtual Reality International Conference, Lacal: France.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Coid, J., Yang, M., Tyrer, P., *et al.* (2006) Prevalence and correlates of personality disorder in Great Britain. *The British Journal of Psychiatry, 188*, 423–431.

Crawford, J.R., & Henry, J.D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *43*, 245-265.

Crick, N.R., & Dodge, K.A. (1996). Social information-processing mechanisms on reactive and proactive aggression. *Personality and Social Psychology Bulletin*, *67*(3), 993-1002.

Des Jarlais DC, Lyles C, Crepaz N. (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *American Journal of Public Health, 94*, 361-366.

DeWall, C.N., & Baumeister, R.F. (2006). Alone but feeling no pain: Effects of social exclusion on physical pain tolerance and pain threshold, affective forecasting, and interpersonal empathy. *Journal of Personality and Social Psychology, 91*, 1-15.

Dolan, M., & Fullam, R. (2004). Theory of mind and mentalizing ability in antisocial personality disorders with and without psychopathy. *Psychological Medicine: A Journal of Research in Psychiatry and the Allied Sciences, 34,* 1093-1102.

EI-Sheikh, M., Harger, J., & Whitson, S. (2001). Exposure to parental conflict and children's adjustment and physical health: The moderating role of vagal tone. *Child Development, 72*, 1617-1636.

Fanselow, M.S., Lester, L.S. (1988). A functional behavioristic approach to aversively motivated behavior: predatory imminence as a determinant of the topography of defensive behavior. In: *Evolution and Learning* (ed. R. C. Bolles and M. D. Beecher), pp. 185–211. Lawrence Erlbaum Associates: Hillsdale, NJ.

Farrington, K. M. (1980). Stress and family violence. In M. A. Straus & G. T. Hotaling (Eds.), *The social causes of husband-to-wife violence* (pp. 94–114). Minneapolis: University of Minnesota Press.

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods.*

Fazel, S., & Danesh, J. (2002). Serious mental disorder in 23,000 prisoners: a systematic review of 62 surveys. *Lancet, 359*, 545–550.

Ferguson, C. J. & Rueda, S. M. (2009). Examining the validity of the Modified Taylor Competitive Reaction Time Test of aggression. *Journal of Experimental Criminology, 5*(2), 121-137.

First, M.B., Gibbon M., Spitzer R.L., Williams, J.B.W., Benjamin L.S. (1997). *Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)*. Washington, D.C.: American Psychiatric Press, Inc.

Gibbon S, Duggan C, Stoffers J, et al. (2010). Psychological interventions for antisocial personality disorder. *Cochrane Database of Systematic Reviews* (6).

Gilligan, C. (1993). *In a Different Voice (3rd ed.)*. Harvard University Press: Cambridge, MA.

Gilligan, J. (1996). Exploring shame in special settings: a psychotherapeutic study. In: C.Cordess & M. Cox (Eds.), *Forensic Psychotherapy: crime, psychodynamics and the offender patient.* (pp. 475-489). Gateshead & Tyne & Wear: Atheneum Press.

Hare, R.D. (2003). *Psychopathy Checklist-Revised technical manual* (2nd ed). Toronto: Multihealth Systems Inc.

Herpertz SC, Werth U, Lukas G, Qunaibi M, Schuerkens A, Kunert H, Freese R, Flesch M, Mueller-Isberner A, Osterheider M, Sass H (2001). Emotion in criminal offenders with psychopathy and borderline personality disorder. *Archives of General Psychiatry, 58*, 737–745.

Hicks, B.M., & Patrick, C. J. (2006). Psychopathy and negative emotionality: Analyses of suppressor effects reveal distinct relations with emotional distress, fearfulness, and anger-hostility. *Journal of Abnormal Psychology, 115,* 276–287.

Lang, P. J., Bradley, M.M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97,* 377–398.

Lang P.J., David M., Öhman, A. (2000). Fear and anxiety: animal models and human cognitive psychophysiology. *Journal of Affective Disorders 61*, 137–159.

Lacey, J.I., Lacey, B.C. (1958). The relationship of resting autonomic activity to motor impulsivity. In The Brain and Human Behavior. *Proceedings of the Association for Research in Nervous and Mental Disease*, pp. 1257–1290. Williams & Wilkins: Baltimore.

Larden, M., Melin, L., Holst, U., & Langstrom, N. (2006). Moral judgement, cognitive distortions and empathy in incarcerated delinquent and community control adolescents. *Psychology, Crime & Law, 12*(5), 453-462.

Levenston, G.K., Patrick, C.J., Bradley, M.M., & Lang, P.J. (2000). The psychopath as observer: Emotion and attention in picture processing. *Journal of Abnormal Psychology, 109,* 373–385.

Linehan, M.M. (1987).Dialectical behavior therapy: A cognitive behavioral approach to parasuicide. *Journal of Personality Disorders, 1*, 328–333.

Lobbestael, J., Arntz, A., Cima, M. & Chakhssi, F. (2009). Effects of induced anger in patients with antisocial personality disorder. *Psychological Medicine*, 39, 557-568.

Lochman, J.E., Wayland, K.K., & White, K.J. (1993). Social goals: Relationship to adolescent adjustment and to social problem solving. *Journal of Abnormal Child Psychology, 21*(2), 135-151.

Loranger, A.W., Sartorius, N., Andreoli, A., Berger, P., Buchheim, P., Channabasavanna, S.M., Coid, B., Dahl, A., Diekstra, R.F.W., Ferguson, B., Jacobsberg, L.B., Mombour, W., Pull, C., Ono, Y. & Regier, D.A. (1994). The International Personality Disorder Examination, IPDE. The WHO/ADAMHA International Pilot Study of Personality Disorders. *Archives of General Psychiatry, 51*, 215-224.

Lorber, M.F. (2004). Psychophysiology of aggression, psychopathy, and conduct problems: a meta-analysis. *Psychological Bulletin*, *130*, 531–552.

Miyahira, S.D., Folen,R.A., Stetz, M., Rizzo, A., & Kawasaki, M.M. (2010). Use of immersive virtual reality for treating anger. *Studies in health technology and informatics*, *154*, 82-6.

Morgan, A. B., & Lilienfeld, O. (2000). A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. *Clinical Psychology Review, 20,* 113–156.

Mühlberger, A., Wiedemann, G., & Pauli, P. (2003). Efficacy of a one-session virtual reality exposure treatment for fear of flying. *Psychotherapy Research, 13*, 323-336.

National Institute for Health and Clinical Excellence (2010). *Antisocial Personality Disorder: Treatment, Management and Prevention [Clinical Guidelines 77]*. London: National Institute for Health and Clinical Excellence.

National institute for Mental Health for England (2003). *Personality Disorder: no Longer a Diagnosis of Exclusion. Policy Implementation Guidance for the Development of Services for People with Personality Disorder*. London: NIMH(E).

National Service User Involvement Network (2011). Unlocking Service User Involvement Practice in Forensic Settings. London: NSUN.

Patrick, C.J., Cuthbert, B.N., & Lang, P.J. (1994). Emotion in the criminal psychopath: Fear image processing. *Journal of Abnormal Psychology*, *3*, 523–534.

Porter, S., Woodworth, M., Earle, J., Drugge, J., & Boer, D. (2003). Characteristics of sexual homicides committed by psychopathic and nonpsychopathic offenders. *Law and Human Behavior*, *27*, 459–470.

Raine, A., Buchsbaum, M., LaCasse, L. (1997). *Brain abnormalities in murderers indicated by positron emission tomography*. Society of Biological Psychiatry

Rizzo, A.A., & Kim, G.J. (2005). A SWOT analysis of the field of VR rehabilitation and therapy, *Presence: Teleoperators and Virtual Environments, 14*(2), 119-146.

Robins, L.N., & Price, R.K. (1991). Adult disorders predicted by childhood conduct problems: results from the NIMH Epidemiological Catchment Area Project. *Psychiatry*, *542*, 116-132.

Robins, L.N., Tipp, J., Pzybeck, T. (1991). Antisocial personality. In: *Psychiatric Disorders in America: The Epidemiological Catchment Area Study* (ed. L. N. Robins and D. A. Regier), pp. 258–290. The Free Press: New York.

Rosenbaum, A., & O'Leary, K. D. (1981). Children: The unintended victims of marital violence. *American Journal of Orthopsychiatry, 51,* 692–699.

Rovira, A., Swapp, D., Spanlang, B., Slater, M. (2009). The use of virtual reality in the study of people's responses to violent incidents. *Frontiers in Behavioral Neuroscience,3*(59), 1-10.

Rummel, R. J. (1970). *Applied Factor Analysis*. Evanston: Northwestern University Press.

Slater, M., A. Antley, A. Davison, D. Swapp, C. Guger, C. Barker, N. Pistrang, and M. V. Sanchez-Vives (2006), "A virtual reprise of the Stanley Milgram obedience experiments", *PLoS ONE* 1 (1).

Sprague, J., & Verona, E. (2010). Emotional conditions disrupt behavioral Control among individuals with dysregulated personality traits. *Journal of Abnormal Psychology, 119,* 409–419.

Swanson, M.C., Bland, R.C., & Newman, S.C. (1994) Epidemiology of psychiatric disorders in Edmonton. Antisocial personality disorders. *Acta Psychiatrica Scandinavica, Supplementum, 376*, 63–70.

Tabachnik, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn and Bacon.

Twenge, J.M., Baumeister, R.F., DeWall, C.N., Ciarocco, N.J., & Bartels, J.M. (2007). Social exclusion decreases prosocial behaviour. *Journal of Personality and Social Psychology, 92*, 56-66.

Urry, H. L., & Gross, J. J. (2010). Emotion regulation in older age. *Current Directions in Psychological Science, 19*, 352-357.

Vaidyanathan, U., Hall, J. R., Bernat, E. M., & Patrick, C. J. (2011). Clarifying the role of defensive reactivity deficits in psychopathy and antisocial personality using startle reflex methodology. *Journal of Abnormal Psychology, 120,* 253–258.

Verona, E., Patrick, C. J., & Lang, A. R. (2002). A direct assessment of the role of state and trait negative emotion in aggressive behavior*. Journal of Abnormal Psychol*ogy, *111*, 249–258.

Verona, E., Sadeh, N., & Curtin, J. (2008). Stress-induced asymmetric frontal brain activity and aggression risk. *Journal of Abnormal Psychology, 118*, 131-145.

Verona, E., Sprague, J., & Sadeh, N. (2012). Inhibitory control and negative emotion processing in psychopathy and antisocial personality disorder. *Journal of Abnormal Psychology*. doi:10.1037/a0025308

Watson, D., Clark, L.A., & Tellegen (1988). Development and validation of a brief measure of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *47*, 1063-1070.

Weber, R. P. (1990). *Basic Content Analysis* (2nd ed.), Newbury Park, CA: Sage.

Webster, C.D., Douglas, K.S., Eaves, D., & Hart, S.D. (1997). *HCR–20: HCR-20: Assessing the Risk for Violence (Version 2).* Vancouver: Mental Health, Law, and Policy Institute, Simon Fraser University.

Part 3: Critical Appraisal

1. Overview

This appraisal briefly outlines my interest in the research area. It then considers the strengths and limitations of using a multi-method approach to measurement and being reliant on technology in psychological research. Next as highlighted in the empirical paper, it gives more detailed consideration to VR as a potential tool to assess and treat risk of aggression and violence in ASPD. Finally, I reflect on some risk management and ethical considerations associated with conducting research with this population.

2. Background interest

My interest in the research area was in part influenced by my experience working clinically in forensic mental health settings, but also informed by my own developmental context. Discussions of 'the mad, bad and dangerous' were commonplace during my childhood, with a number of family members working in the legal system. I was curious about why some people ended up in the criminal justice system and, being raised in inner city London, I could not fail to notice how crime and mental health seemed to systematically vary with social disadvantage and inequality. These early experiences formed the basis for my developing interest in ASPD.

Before training I worked in a number of secure hospitals with diverse clinical presentations and varying levels of criminogenic risk. I was particularly struck by how individuals diagnosed with ASPD presented apparent paradoxes, on the one hand displaying prosocial behaviour and a capacity for self-reflection, and on the other engaging in challenging behaviour against an extensive backdrop of criminality. Rarely did I encounter the 'cold-blooded and calculated' criminal stereotype. My observations were instead of individuals who had profound difficulty inhibiting their responses to painful emotions, in particular shame, humiliation and

rejection. These observations were captured in explanatory models of ASPD in which violence and humiliation are conceptualised as strategies for coping with the impact of shame on an inflated but fragile sense of self (Gilligan, 1996).

When Peter Fonagy presented the opportunity to conduct research with this clinical group, using an exciting new approach to prime emotional arousal, it was the exact opportunity I was seeking. Intuitively it made sense that like other Cluster B disorders, ASPD could be partly understood as a disorder of emotion regulation leading to maladaptive behaviour in the context of interpersonal distress. When I began researching this topic I was therefore surprised to find a lack of empirical research involving individuals with a well-defined diagnosis of ASPD, and only one study that had directly sought to measure emotional reactivity (Lobbestael, Arntz, Cima & Chakhssi, 2009).

3. Measurement issues

3.1 Multi-method measurement

From the outset there were several key objectives to take into consideration when designing an approach to measurement. An initial priority was to capture the full associative network comprising emotions (Berkowitz, 1990) without burdening participants with an exhaustive list of tasks. Second, given the exploratory nature of the study both in terms of working with an under-researched population and using VR provocation, it was important to elicit participants' idiosyncratic experiential perspectives. This needed to be balanced alongside valid indices of emotionality that could reliably detect change, and with establishing a meaningful strategy for coding and interpreting observational data. Consequently a multi-method measurement approach was adopted. This entailed several strengths and drawbacks on a theoretical, practical and interpretive level.

A conceptual strength of the measurement approach is that multi-method research can be less vulnerable to inaccurate inferences arising from error because it enables cross-validation and cross-fertilization of research procedures and findings (Brewer & Hunter 1989). When triangulated outcomes have good convergent validity (providing they are not vulnerable to the same sources of measurement error as one-another), conclusions can be made with greater conviction (Bryman, 2001). When disparity is observed between outcomes, consideration of possible explanations for the clash is necessitated (e.g. Deacon, Bryman & Fenton, 1998).

In terms of the current study, convergent validity was mixed, with indices of positive emotionality correlating with conciliatory behaviour but the relationship between negative cognitions and aggression falling short of significance. This encouraged me to think critically about sources of measurement error (e.g. a time-lag measurement effect confounder) and hypothesise about factors that may have influenced the relationship between outcomes. For example, I tentatively suggested that through treatment, participants with ASPD had learnt to inhibit behavioural manifestations of aggression but remained prone to anger-oriented thinking. This highlighted fresh avenues for research, which would have been missed if a mono-method measurement approach had been employed.

A further strength of the measurement approach was that it avoided heavy reliance on self-report questionnaires. I considered this particularly important in research with participants diagnosed with ASPD for several reasons. First, at the study's outset I was aware of research suggesting that because lying is a central characteristic of ASPD, ratings on self-report tools may reflect text-appropriate ratings, rendering them unreliable indicators of internal states (Herpentz et al., 2001). Second, educational disadvantage (and related literacy problems) is a reliable concomitant of ASPD (Swanson, Bland & Newman, 1994), reflected in the low levels of academic qualifications observed in the present study. I hoped that by

restricting the number of tasks that required reading and writing, and offering a high level of support with questionnaires when required, the research burden on clinical participants would be reduced. This seemed to be effective as many participants reported enjoying the experience of taking part, offering their time in the event of future studies.

Third, my personal observation of working clinically with this client group is that structured learning environments can provoke unhappy memories of school, arousing anxiety about failure and negative evaluation. I was keen to avoid this, primarily for ethical reasons but also as a potential confounder in emotional reactivity outcomes. I was also aware that forensic service-users are often required to complete extensive self-report batteries about their thoughts and feelings, and I was concerned that this would engender more frustration amongst clinical participants, thereby introducing an additional confounder in reactivity outcomes. Therefore balancing structured measures, interviews and experiential tasks seemed most appropriate. Nevertheless, it remained possible that the research process had a differential impact on groups, by virtue of clinical participants being understandably concerned about how they would be perceived in light of the stigma associated with having a criminal history and a diagnosed personality disorder.

The main drawback associated with a multi-method approach was that with so many sources of data to analyse, it was not possible to do each component justice within the confines of the project. For example it would have been interesting to analyse change scores on individual PANAS (Watson, Clark & Tellegen, 1988) items but the time constraints coupled with the risk of increasing the Type I error rate with additional statistical tests, meant that this was not possible. Similarly, an interview protocol was developed with the intention of gaining a more in depth picture of participants' experiences than would be possible with a standardised tool. However, the decision to only analyse questions pertaining to cognitions (in light of

time constraints), resulted in a somewhat reductionist approach and meant that some of the richness of the data was lost.

For example, many clinical participants described how the scenario elicited similar emotions to those they had experienced in fights many years prior, making them aware of their continued vulnerability to respond with aggression and reinforcing the importance of avoiding places where they could encounter conflict (e.g. pubs) in the future. A number of clinical participants with comorbid substance misuse diagnoses further commented on how seeing virtual pint glasses triggered dormant cravings. These qualitative reports made me think about the potential for VR-based treatments for offenders with ASPD (see below). It was disappointing not to be able to capture these reflections in the results of the empirical paper, and in retrospect it would have been useful to fully exploit experiential interview reports and administer a standardised measure of cognitions as an adjunct. Nevertheless, the three 'cognitive salience' constructs that emerged from the data could contribute to future research by forming the basis of a structured tool in the event of replication.

3.2 Technological considerations

Several lessons were learnt from conducting a project in which the success of data collection hinged on technological equipment. This was most notable with regard to physiological data, which regrettably proved unusable. From the point at which the project was first conceptualised, heart-rate was considered a key indicator of arousal, and muscle tension (triceps) an objective gauge of participants' desire to respond impulsively with physical aggression. Accordingly, much of the preparation time prior to data collection involved familiarising myself with the physiological equipment (NEXUS 4) and learning how to interpret outcomes.

It quickly became apparent that there was an absence of departmental expertise available to inform this process and therefore much of my learning

occurred during testing with participants. Unfortunately my limited experience meant that I did not detect subtle anomalies in the data arising from an equipment fault, which were probably exacerbated by interference from the virtual reality equipment. It was only after completion of testing that I managed to consult with suitably qualified researchers and I was informed that the data was not salvageable. This was disappointing as physiological data would have not only been useful for hypothesis testing but also for validating the anger-arousal potential of the VR scenario.

On reflection it was over-ambitious to expect to master physiological measurement techniques alongside the significant challenges associated with recruiting a hard-to-reach sample and becoming accustomed to the nuances of virtual reality. In the event of replication, it will be important to ascertain if experts are available to assist from the outset and thereafter allocate sufficient time to conduct a pilot in which equipment faults can be detected and trouble-shooted.

Using state-of-the-art VR technology also incurred certain advantages and disadvantages. On the one hand it enabled me to capture and analyse 'live' behaviour during ecologically valid conditions of provocation. This represented an important step in better understanding emotional and behavioural reactivity in an under-researched clinical population. However, during both the pilot phase and testing with healthy volunteers, there were several occasions when I had to cancel participants at the last minute due to equipment failures.

A further limitation was that the VR equipment was not sophisticated enough for avatars to respond to participants once the altercation had begun. A number of participants commented that the realism and believability of VR was compromised by the fact that they could not effect change in the outcome of the scenario. In addition to impacting on 'presence', this raised the possibility that when participants attempted to intervene, an absence of reinforcing social cues led them to 'give up'

more easily, thereby reducing the authenticity of their behaviour. Additionally, although in many ways the VR scenario was highly realistic, it occasionally appeared clunky. For example in the initial conversation with the victim avatar, participants sometimes misheard his responses and requested clarification. The system was not sophisticated enough to repeat small extracts from the pre-recorded array of utterances, and my observations were that this temporarily suspended participants' sense of immersion.

The scenario content also introduced unanticipated difficulties. During recruitment several clinical participants refused to take part as they had a strong existing affiliation with a football team and were understandably unwilling to take on the role of an Arsenal supporter. This made recruitment of an already hard to access sample more difficult. Furthermore a number of participants from both clinical and healthy volunteer groups commented in interview that they felt uncomfortable during their initial conversation with the victim avatar as they did not know the answers to his questions about football. Future studies may benefit from using an alternative scenario that elicits less apprehension in participants during both the recruitment and participation phases.

4. Virtual risk assessment and treatment

4.1. Virtual risk assessment

The findings from the empirical study raise the possibility that VR could be used to assist the risk assessment process in forensic personality disorder services. At a population level there is a strong statistical association between the diagnosis of ASPD and violent offending behaviour (Singleton, Melzer & Gatwood, 1998). However, ASPD is a very broad diagnostic category (DSM-IV; APA, 2000), encompassing people who never commit offences as well as those who commit the most serious crimes. Consequently, the diagnosis alone is of little value as an indicator of risk (NICE CG 77, 2009). The assessment of risk of violence in ASPD is

further complicated by the fact that some patients may be persuasive and deceptive, rendering the clinical interview an inaccurate guide to the severity of the disorder and associated risks.

It is generally accepted that the best way to assess violence risk is through structured clinical judgement (Monahan, Steadman & Silver et al., 2001), namely based on standardised information alongside prior knowledge about an individual case. Several measures have been developed to facilitate this process (e.g. the HCR-20; Webster, Douglas, Eaves & Hart, 1997), with the key aim of balancing specificity and sensitivity in evaluations of probability. Although these instruments are moderately predictive of recidivism (e.g. Dahle, 2006), they have limited accuracy predicting behaviour at an individual level in clinical practice (NICE GC 77, 2009). As a result, the development of any measure that discriminates between the degree and severity of ASPD is likely to assist the process of risk assessment (NICE CG 77, 2009).

It is possible that VR provocation could inform this process by balancing judgement based on general patterns of behaviour, whilst providing idiosyncratic information about individual clients' personal risk factors. Specifically, in the event of large-scale data collection, it may be possible to develop an algorithm based on the four markers of affect, cognitions, physiological arousal and behaviour during VR. This would enable an individual's pattern of reactivity to be compared to normative samples, such that their probability of reoffending could be calculated with sufficient specificity and sensitivity. At the same time, there may be scope for clinicians to use their existing knowledge of individual clients to build VR scenarios that reflect idiosyncratic high risk situations, allowing more precise evaluation of that particular individual's risk of recidivism. Predictive validity could then be evaluated by obtaining longitudinal follow-up data of actual rates of reoffending.

Evidently, developments in VR risk assessment would necessitate large scale normative data from which reactivity cut-off values could be derived.

Furthermore the scope for clinicians to develop individual scenarios will depend on the cost and flexibility of VR equipment in the future. Consequently the potential for VR-based risk assessment requires ongoing research, using the most sophisticated technology available.

4.2. Virtual anger treatment

If, as the results of the empirical paper suggest, aggression in ASPD primarily reflects poor behavioural inhibition during anger-arousing circumstances, then aggression in this population may be amenable to exposure therapy. Exposure is a central ingredient of CBT and is thought to be effective by way of de-conditioning the anger learning cycle (Foa & Rothbaum, 1998), and facilitating the integration of corrective information (Foa & Hearst-Ikeda, 1996). However traditional methods of anger exposure are subject to several limitations. Imaginal techniques are dependent on an individual's capacity to visualise stimuli, role-play may lack sufficient realism to provoke an emotional response, and in vivo exposure raises serious ethical and safety implications.

For these reasons, several VR-based anger exposure protocols have been developed and there is preliminary evidence to support their anger-arousal potential (e.g. Miyahira et al., 2010), echoed by the results of the current empirical study. When viewed alongside the fact that CBT offence-focused treatment is commonplace in the UK (e.g. Ireland, 2004) and shows preliminary efficacy with ASPD populations (Gibbon et al., 2010), it is plausible that VR exposure could offer an effective new treatment component for violent offenders with ASPD.

In addition to the foregoing, VR exposure may lend itself ideally to progress monitoring in anger treatment, overcoming the limitations associated with traditional evaluative methods. Historically, assessment of skills acquisition has been largely limited to self-ratings and/or external ratings (e.g. staff observations), which are vulnerable to social desirability and stereotyping biases (Frankfort-Nachmias & Nachmias, 2000). Although role-play exercises may be used to gauge the

development of anger-management skills, they may lack generalisability and be difficult to implement with large numbers of service-users. By contrast, VR exposure can be administered relatively easily and may provide a more objective, ecologically-valid gauge of treatment progress.

Evidently, a key factor in developing 'virtuous reality' treatment protocols for ASPD will be to design scenarios that can sufficiently prime anger and evoke realistic responses. This will necessitate a cooperative interplay between software engineers, health care providers and service-users, representing an exciting opportunity for collaborative research initiatives in the future.

5. Final reflections

The research process raised several ethical and risk management considerations that warrant discussion. First, owing to the anger-arousal potential of the scenario, participants' violent offending histories, and the absence of prior research to draw upon, I experienced some anxiety and apprehension about the potential for violence arising from participation in the study. These concerns were echoed by the ethics committee and by clinicians assisting in the recruitment process. However, I also felt that precisely because research with this population was lacking, it was important to go ahead with the study in order to facilitate a better understanding of the processes implicated in ASPD aggression. For me this was particularly important in light of the general paucity of evidence to support efficacious interventions for ASPD (Gibbon, Duggan, Stoffers, et al., 2010).

I was also mindful of the ethical obligation to avoid pejorative stereotypes, whilst adopting a sensible approach to risk management. I found it useful to remember that clinical participants had been judged by the Ministry of Justice and their respective clinical teams to be sufficiently safe to be living in – or spending extended periods of time in – the community. At the same time it was essential to

implement a rigorous risk management protocol, including a no-lone-working policy, carrying a personal alarm and including strategies for reducing participants' residual arousal following VR (e.g. a relaxation exercise).

As noted, at the outset of the project I was aware of research suggesting that self-report tools with ASPD samples may be an inaccurate gauge of internal states. However, as the project unfolded I saw this is less problematic because participants demonstrated a willingness to discuss their propensity towards aggression. For me this illustrates the importance of holding clinical characteristics in mind when designing and conducting research, whilst remaining flexible to individual variation and being careful not to apply negative stereotypes. This may be particularly important when collaborating with participants diagnosed with ASPD, given that they are likely to have experienced a high level of stigma and discrimination in the past (NICE CG 77, 2009).

Having worked clinically with service-users diagnosed with ASPD, I am personally invested in developing the rehabilitative pathway, which necessitates evidence of improved psychological wellbeing and reduction in risk. However, like any psychological condition, successful treatment hinges on developing a coherent understanding of the mechanisms that give rise to and maintain problems. It is unfortunate that historically this group of offenders has shared the 'untreatability' stigma associated with psychopathy, without the elusive glamour. I hope that my reflections on the research process will encourage future researchers to recognise that with sufficient planning and determination, the barriers to conducting research with this population can, and should, be overcome.

REFERENCES

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders, (Revised 4th ed.)*. Washington DC: APA.

Berkowitz, L. (1990). On the formation and regulation of anger and aggression: A cognitive-neoassociationistic analysis. *American Psychologist*, 45, 494–503.

Brewer, J. & Hunter, A. (1989) *Multimethod research: A synthesis of styles*. Newbury Park, CA: Sage.

Bryman, A. (2001). *Social Research Methods*. Oxford, UK: Oxford University Press.

Dahle, K.P. (2006). Strengths and limitations of actuarial prediction of criminal reoffence in a German prison sample: a comparative study of LSI-R, HCR-20 and PCL-R. *International Journal of Law and Psychiatry, 29*, 431–442.

Deacon, D., Bryman, A., & Fenton, N. (1998). Collision or collusion? A discussion of the unplanned triangulation of quantitative and qualitative research methods. *International Journal of Social Research Methodology,* 1, 47-63.

Foa, E.B., & Hearst-Ikeda, D. (1996). Emotional dissociation in response to trauma: an information-processing approach. In: *Handbook of Dissociation: Theoretical, Empirical, and Clinical Perspectives* (eds. L. K. Michelson & W. J. Ray)*,* pp. 207 -224. New York: Plenum Press.

Foa, E.B., & Rothbaum, B.A. (1998). Treating the trauma of rape: *Cognitive behavioral therapy for PTSD*. New York: Guilford Press.

Frankfort-Nachmias, C., & Nachmias, D. (2000). *Research Methods in the Social Sciences, (6th edn*.) Worth: New York.

Gibbon S, Duggan C, Stoffers J, et al. (2010). Psychological interventions for antisocial personality disorder. *Cochrane Database Syst Rev*.;(6):CD007668.

Gilligan, J. (1996). Exploring shame in special settings: a psychotherapeutic study. In: C.Cordess & M. Cox (Eds.), *Forensic Psychotherapy: crime, psychodynamics and the offender patient.* (pp. 475-489). Gateshead & Tyne & Wear: Atheneum Press.

Herpertz, S.C., Werth, U., Lukas, G., Qunaibi, M., Schuerkens, A., Kunert, H., Freese, R., Flesch, M., Mueller-Isberner, A., Osterheider, M., Sass, H. (2001). Emotion in criminal offenders with psychopathy and borderline personality disorder. *Archives of General Psychiatry*, 58, 737–745.

Ireland, J.L. (2004). Anger management therapy with male young offenders: An evaluation of treatment outcome. *Aggressive Behavior*, 30, 174-175.

Lobbestael, J., Arntz, A., Cima, M. & Chakhssi, F. (2009). Effects of induced anger in patients with antisocial personality disorder. *Psychological Medicine*, 39, 557-568.

Miyahira, S.D., Folen, R.A., Stetz, M., Rizzo, A., & Kawasaki, M.M. (2010). Use of immersive virtual reality for treating anger. *Studies in health technology and informatics*, 154, 82-6.

Monahan, J., Steadman, H. J., Silver, E., *et al.* (2001). *Rethinking Risk Assessment. The MacArthur Study of Mental Disorder and Violence.* Oxford: Oxford University Press.

National Institute for Health and Clinical Excellence (2010). *Antisocial Personality Disorder: Treatment, Management and Prevention* [Clinical Guidelines 77]. London: National Institute for Health and Clinical Excellence.

Singleton, N., Melzer, H., Gatward, R., et al. (1998). *Psychiatric Morbidity Among Prisoners in England and Wales*. London: The Stationary Office.

Swanson, M.C., Bland, R.C., & Newman, S.C. (1994). Epidemiology of psychiatric disorders in Edmonton. Antisocial personality disorders. *Acta Psychiatrica Scandinavica, Supplementum, 376*, 63–70.

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of a brief measure of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 47, 1063-1070.

Webster, C.D., Douglas, K.S., Eaves, D., & Hart, S.D. (1997). *HCR–20: HCR-20: Assessing the Risk for Violence* (Version 2). Vancouver: Mental Health, Law, and Policy Institute, Simon Fraser University.

APPENDICES

APPENDIX 2: RECRUITMENT AIDES

Advertisement for healthy volunteers

Research Poster for Healthy Volunteers

Version 1: 16.07.2011

This project has been approved by UCL ethics committee

VIRTUAL REALITY STUDY

We are looking for men to take part in a doctoral research project looking at how people respond to virtual reality scenarios. The study will take up to 90 minutes; you will be paid for your time.

For more information please email ophelia.phillips.09@ucl.ac.uk

Clinician Letter Requesting Recruitment Support
Version 1: 16.07.2011
Project approved by Camberwell & St Giles Research Ethics Committee

(To appear on Trust headed paper)

Ophelia Phillips
Research Department of Clinical,
Educational and Health Psychology
University College London
1-19 Torrington Place
London WC1E 6BT
Email: ophelia.phillips.09@ucl.ac.uk

Date

Clinician name & address

Dear Dr.

We are writing to request your support in helping to identify and recommend potential participants for a research project. This study is investigating the way in which people respond when they witness aggressive interactions in a virtual reality environment. The research does not involve any medical procedures. It is intended to contribute to an understanding of psychophysiological markers of antisocial personality disorder.

We are looking to recruit 20 participants with a diagnosis of antisocial personality disorder who are currently in receipt of care from inpatient and outpatient NHS services. Potential participants are required to be English-speaking and possess basic literacy skills. Due to the use of virtual reality, patients with an epileptic condition will not be able to take part. Additionally, patients with an active major mental illness will not be eligible to participate.

The study will take place in the Department of Computer Science, University College London (UCL). Therefore, potential participants who currently reside in inpatient services are required to have been granted unescorted community leave of a sufficient time period to travel to and from UCL and take part in the research. It is estimated that participation will take 1 hour 30 minutes (excluding travel time). Participants will be paid for their involvement in the study.

We include our information sheet for further details about the study, which you may wish to discuss with potential participants. If your patient expresses interest in participating, Ophelia Phillips (researcher) will come to meet with them at a convenient time to explain the research in more detail and obtain informed consent. In this instance, Ophelia Phillips will liaise with the patient and yourself to arrange an appointment at UCL.

We do not anticipate any adverse effects from taking part in this study. However, if for any reason concerns are raised regarding your patient's safety or wellbeing during their participation, a member of the research team will contact you (or a pre-agreed designated member of their clinical team) directly to inform you of this.

Please let us know if you have any questions.

Yours sincerely,

Ophelia Phillips
Trainee Clinical Psychologist
University College London / Camden & Islington NHS Foundation Trust

Professor Peter Fonagy
Head of the Department of Clinical, Educational & Health Psychology
University College London

Dr. Chris Barker
Reader in Clinical Psychology
University College London, Department of Clinical, Educational & Health Psychology

Dr. Stephen Butler
Senior Lecturer, Department of Clinical, Educational & Health Psychology
University College London

Responsible Clinician letter about participation:
Version 1: 16.07.2011
Project approved by Camberwell & St Giles Research Ethics Committee

(To appear on Trust headed paper)

Ophelia Phillips
Research Department of Clinical,
Educational and Health Psychology
University College London
1-19 Torrington Place
London WC1E 6BT
Email: ophelia.phillips.09@ucl.ac.uk

Date
Clinician name & address

Dear Dr.

re. patient name
d.o.b.

We are writing to inform you that your patient has agreed to take part in a research project, investigating the way in which people respond when they witness aggressive interactions in a virtual reality environment.

The research does not involve any medical procedures. It is intended to contribute to an understanding of psychophysiological markers of antisocial personality disorder. We do not expect there to be any adverse effects, but we include our information sheet in case your patient wishes discuss any aspects of the study with you.

In addition, if any concerns about your patient's safety or wellbeing are raised during their participation, a member of the research team will contact you (or a pre-agreed designated member of their clinical team) to inform you of this.

Please let us know if you have any questions.

Yours sincerely,

Ophelia Phillips
Trainee Clinical Psychologist
University College London / Camden & Islington NHS Foundation Trust

Professor Peter Fonagy
Head of the Department of Clinical, Educational & Health Psychology
University College London

Dr. Chris Barker
Reader in Clinical Psychology
University College London, Department of Clinical, Educational & Health Psychology

Dr. Stephen Butler
Senior Lecturer, Department of Clinical, Educational & Health Psychology
University College London

APPENDIX 3: ETHICAL APPROVAL LETTERS

UCL ETHICAL APPROVAL
UCL GRADUATE SCHOOL
UCL RESEARCH ETHICS COMMITTEE

Professor Mel Slater
Department of Computer Science
UCL

9 June 2011

Dear Professor Slater

Re: Ethics Application: 0206/002: Bystander responses to violent emergencies – a virtual reality experiment

I am pleased to confirm that I have approved your request to extend the above project for a further 18 months; until June 2013 and to include Ophelia Phillips and her supervisor, Dr Chris Barker, as additional researchers.

I also approve your proposal to widen the recruitment criteria to men over 18, to use additional, non-intrusive standardised questionnaire depending on the results of the pilot work and to use heart rate and EMG recordings via a non-invasive physiological measurement device.

As always, please notify the Administrator of the Ethics Committee, Helen Dougal, if you propose to make any further amendments. It is also your responsibility to report to the Committee any unanticipated problems or adverse events involving risks to participants or others.

On completion of the research you must submit a brief report (a maximum of two sides of A4) of your findings/concluding comments to the Committee, which includes in particular issues relating to the ethical implications of the research.

With best wishes

Yours sincerely

Sir John Birch
Chair, UCL Research Ethics Committee

NRES ETHICS APPROVAL

FILE COPY

NRES Committee London - Camberwell St Giles
(Formerly known as The Joint South London and Maudsley and Institute of Psychiatry Research Ethics Committee)
Administrative address: Victoria House
Capital Park
Fulbourn
Cambridge
CB21 5XB

Telephone: 01223 597509
Facsimile: 01223 597645

05 October 2011

Dr Stephen Butler
Senior Lecturer
University College London
Research Department of Clinical, Educational and Health Psychology, University College London
4th Floor, 1-19 Torrington Place,
London
WC1E 7HB

Dear Dr Butler

Study title: **An investigation into the emotional processing of aggressive virtual reality stimuli in individuals with a diagnosis of antisocial personality disorder**
REC ref: **11/LO/1276**

Thank you for your letter of 14 September 2011, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair, in consultation with Mrs Sally Gordon Boyd.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Ethical review of research sites
NHS sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Non-NHS sites

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Advertisement: Poster for healthy volunteers: 1		16 July 2011
Covering Letter		16 July 2011
Evidence of insurance or indemnity: University College London Hospitals		18 July 2011
GP/Consultant Information Sheets	1	16 July 2011
Investigator CV: Stephen Butler		
Letter of invitation to participant: Recruitment advertisement for healthy volunteers:1		16 July 2011
Other: Letter to GP requesting recruitment support	1	16 July 2011
Other: Script extract from virtual reality scenario	1	16 July 2011
Other: Progressive muscular relaxation script	1	16 July 2011
Other: CV for academic supervisor: Peter Fonagy		
Other: CV for student: Ophelia Phillips		16 July 2011
Other: CV for academic supervisor : Dr Chris Barker		
Other: Vignette - shaking hands and making up	1	14 Sept 2011
Participant Consent Form: Clinical participants	4	20 July 2011
Participant Consent Form: Healthy volunteer participants	4	20 July 2011
Participant Information Sheet: Clinical participants	5	14 Sept 2011
Participant Information Sheet: Healthy volunteer participants	5	14 Sept 2011
Protocol	3	20 July 2011
Questionnaire: STAXI-2		
Questionnaire: PANAS		
Questionnaire: Multidimensional Personality Questionnaire		
REC application: 79331/235184/1/348		26 July 2011
Referees or other scientific critique report: Research proposal review form		15 March 2011

Response to Request for Further Information: from Dr Stephen Butler		14 Sept 2011
Summary/Synopsis: Flowchart: 2		20 July 2011

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document *"After ethical review – guidance for researchers"* gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

Feedback

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

Further information is available at National Research Ethics Service website > After Review

11/LO/1276	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project

Yours sincerely

Mr John Richardson
Chair

Email: charis.bailey@eoe.nhs.uk

Enc: "After ethical review – guidance for researchers" *[SL-AR2]*

Copy: Dave Wilson
University College London
R&D Dept., 1st Floor, Maple House
Rosenheim Wing
25 Grafton Way
London
WC1E 6DB

Ms Angela Williams,
North Central London Research Consortium
R&D Office,
3rd Floor
West Wing, St Pancras Hospital
4 St Pancras Way
London
NW1 0PE

Mrs Ophelia Phillips
Research Department of Clinical, Educational and Health Psychology,
University College London
4th Floor, 1-19 Torrington Place,
London
WC1E 7HB

APPENDIX 4: INFORMATION FOR PARTICIPANTS

Information sheet for clinical participants

Version 5: 14.09.2011

PROJECT APPROVED BY NRES COMMITTEE LONDON – CAMBERWELL ST GILES

(To appear on Trust headed paper)

Information Sheet: Clinical Participants

You will be given a copy of this information sheet.

PROJECT TITLE: A VIRTUAL REALITY STUDY OF HOW PEOPLE RESPOND WHEN WITNESSING AGGRESSIVE INTERACTIONS

Researcher Name	Ophelia Phillips
Work Address	Department of Clinical, Educational and Health Psychology, University College London, 1-19 Torrington Place, 4th Floor, London WC1E 7HB
Contact Details	email: ophelia.phillips.09@ucl.ac.uk

INVITATION TO TAKE PART

We would like to invite you to take part in our research study. Before you decide we would like you to understand why the research is being done and what it would involve for you. **One of our team will go through the information sheet with you and answer any questions you have.** We'd suggest this should take about 10 minutes. (Part 1 tells you the purpose of this study and what will happen to you if you take part. Part 2 gives you more detailed information about the conduct of the study). Ask us if there is anything that is not clear. Talk to others about the study if you wish.

PART 1: DETAILS OF STUDY

Purpose: This research project is looking at people's reactions to virtual environments. This takes the form of cutting edge technology, where a room is filled with 3-Dimensional projections that simulate a real life scenario with moving interactive characters. The aim of the study is to develop our understanding of how people respond when they witness aggressive interactions between other people. We are interested in how people respond physiologically, emotionally and behaviourally. This project is part of a student research project. Please take time to read the following information carefully and ask us if there is anything unclear or if you would like more information. If you are happy to take part in the study you will be asked to sign an informed consent form.

Why have I been invited to take part in this study? You have been invited to take

part in this study because we are looking for males aged between 18 and 60 years who have contact with forensic mental health services and have been given a diagnosis of Antisocial Personality Disorder. We are hoping to recruit twenty participants. The study will take place in the Department of Computer Science, University College London. It is therefore only possible for you to take part if you have been granted unescorted community leave of sufficient time period (1 hour 30 minutes plus travel to and from University College London) by the Ministry of Justice, and that this is supported by your clinical team.

Do I have to take part? No, it is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep, and be asked to sign a consent form. You are still free to withdraw at any time, and you do not need to give a reason for this. This would not affect the standard of care you receive.

What will happen if I decide to take part? In addition to your direct involvement in the study, if you agree to take part one of the researchers Ophelia Phillips, Trainee Clinical Psychologist will have access to your medical records. This is in order to find out some basic information about you (e.g. your age and diagnosis), how long you have been in contact with forensic mental health services, and what your criminal history is. All of this information will be stored in a way to ensure that you are not identifiable. Your direct participation in the project involves five phases and is likely to take between an hour and an hour-and-a-half in total.

Phase 1 - Questionnaires: Before entering the virtual environment, we invite people to complete three questionnaires which ask about your thoughts, feelings and behaviour in general, as well as how you feel at the time of the study. There are no right or wrong answers. One of these questionnaires will be given both before and immediately after the virtual environment experience. Completion of questionnaires will take approximately 45 minutes and you will be able to ask for assistance from a member of the research team if you require.

Phase 2 - Physiological measures: We invite people to have their physiological responses measured in two ways; muscle tension and heart rate. To do this we will request to attach two electrodes to you; one to your upper arm and one to your chest. This will be carried out by a male member of the research team who has experience of doing so, in line with standard health and safety conventions. Another member of the research team will also be present during electrode placement for further assistance.

Once this equipment has been attached, we ask participants to enter the virtual reality room and explore the environment that depicts a pub for two minutes. You will be given instructions in the use of virtual reality before entering. Once you are familiar with the use of virtual reality, you will be asked to put on glasses that produce three-dimensional images. During this initial two minute phase, a member of the research team seated outside the virtual reality room will take some measurements of your resting heart-rate and muscle tension using specialist computer software.

Phase 3 - Virtual Reality: After Part 2 has been completed, you will be invited to remain in the virtual pub and continue observing the environment. The scenario will last for two minutes and twenty seconds during which you will encounter and be invited to interact with two virtual characters. You will witness an interaction between these two characters; the language used will be realistic to the scenario. You will also be asked to respond to the situations and characters to you encounter as you would in everyday life. You will also be asked to wear an Arsenal football

team scarf, and to imagine that you are an avid Arsenal supporter.

During your time in the environment a member of the research team will take further muscle tension and heart rate measurements from outside the virtual reality room. **Your behaviour and verbal responses in the virtual environment will be video-recorded** for the purpose of analysing data from the study. Once this has been coded by the researcher, the video tapes will be wiped clean.

There will be another researcher directly outside the virtual room at all times to ensure you feel comfortable during the exercise.

Phase 4 – **Post-questionnaire and Interview**: Once you have completed the exercise you will be asked to complete one of the written questionnaires from Stage 1 a second time. You will then be interviewed briefly about the thoughts and feelings you experienced during virtual reality. Phase 4 is estimated to take 10 minutes. Any data obtained from the interview will be coded. Direct quotes from your interview will NOT be used in the write-up of this study.

Phase 5 – Relaxation and scenario ending: You will be asked to take part in a five minute relaxation exercise, which will be read out by one of the researchers. You will then be read a vignette by one of the researchers, explaining how the scenario ended between the two characters in the bar.

Expenses and payment: To thank you for your time, you will be paid £6.00 per 30 minutes of your time spent, which is likely to be 1 hour 30 minutes (totalling £18.00). You will also be reimbursed for travel expenses to and from University College London.

What are the possible disadvantages and risks of taking part?
Information about the virtual reality equipment: When people use virtual reality systems they occasionally experience a degree of nausea. If at any time you wish to stop taking part in the study for this or any other reason, please say so and we will stop.

There is some research to suggest that people using virtual reality might experience some disturbance in vision afterwards. No long term studies are known to us, but some studies which have conducted testing after about 30 minutes, have found that the effect is sometimes still there. It is therefore advised that you do not drive a car, motorcycle or operate complicated machinery in the four hours flowing virtual reality. With all kinds of video equipment there is a possibility that an epileptic episode may be generated. **If you have epilepsy please tell us. We would not want you to take part in the study in this case.**

Information about the virtual scenario: It is possible that you may experience a degree of stress or apprehension when witnessing aggression escalating between virtual characters. If at any time you wish to stop taking part for this or any other reason, please tell us and we will stop.

Some researchers have expressed concern that witnessing violence and aggression in virtual reality may increase the likelihood of people becoming aggressive after being exposed to these virtual scenarios. To our knowledge no research has directly investigated this. However, recent research into violent video-games and increased aggression in players has found no evidence to support this concern. If however you do notice an increase in aggressive thoughts or urges following your time in the virtual environment and after the relaxation exercise, please tell us. In this case we will talk to you about your concerns in order to offer you further support.

What are the benefits of taking part? We cannot promise the study will help you but the information we get from this study will help improve the understanding of Antisocial Personality Disorder. Additionally, some people report that they enjoy taking part in virtual reality, finding it interesting and novel.

What if there is a problem? Any complaint about the way you have been dealt with during the clinical trial or any possible harm you might suffer will be addressed. Detailed information concerning this is given in Part 2 of this information sheet.

Will my taking part in the study be kept confidential? All information from the study will be kept confidential and you will not be identified. As participation is anonymous, it will not be possible for us to withdraw your data once you have returned your questionnaires.

PART 2: ADDITIONAL INFORMATION

What will happen if I don't want to carry on with the study? If you withdraw from the study, we will destroy all your identifiable information, but we may use non-identifiable data that we have collected up until your withdrawal.

What if there is problem? Every care will be taken in the course of this study. However, in the unlikely event that you are injured by taking part, compensation may be available.

If you suspect that the injury is the result of the Sponsor's (University College London) or the hospital's negligence then University College London is insured for this study. After discussing with Ophelia Phillips, please contact Dr Stephen Butler (telephone number: 0207 679 5985), who is the Chief Investigator for the research and is based at the Department of Clinical, Educational & Health Psychology, University College London.

If you have a concern about any aspect of this study, you should ask to speak to the researcher (Ophelia Phillips) who will do her best to answer your questions (telephone number 0207 916 9189). If you remain unhappy and wish to complain formally about any aspect of the way you have been approached or treated by members of staff or about any adverse events you may have experienced due to your participation in the research, the normal National Health Service complaints mechanisms are available to you. Please ask the researcher, Ophelia Phillips if you would like more information on this. Details can also be obtained from the Department of Health website: http://www.dh.gov.uk

Will my taking part in this study be kept confidential? All information which is collected about you during the course of the research will be kept strictly confidential, and any information about you which leaves the hospital/surgery will have your name and address removed so that you cannot be recognised. All data concerning you will be stored in a locked facility.

Your Responsible Clinician will be informed of your involvement in this research, but will not be provided with information about the data we obtain from you such as your responses to questionnaires. All data gathered will be kept strictly within the research team. If the researcher has any concerns about your health and safety during your participation in the study, they will notify your Responsible Clinician of this.

What will happen to the results of the research study? The results of the research will be analysed as part of a doctorate in clinical psychology and the findings are likely to be published in a scientific journal. Participants will not be identified in any report or publication. Please inform Ophelia Phillips if you would

like to be sent a copy of the study's findings.

Who is organising and funding this research? University College London are sponsoring and funding this research.

Who has reviewed this study? All research in the NHS is looked at by independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by Camberwell & St Giles Research Ethics Committee.

All data will be collected and stored in accordance with the Data Protection Act 1998.

Information Sheet for Healthy Volunteers

Version 5: 14.09.2011

PROJECT APPROVED BY UCL AND NRES COMMITTEE LONDON – CAMBERWELL ST GILES

Information Sheet: Healthy Volunteer Participants

You will be given a copy of this information sheet.

PROJECT TITLE: A VIRTUAL REALITY STUDY OF HOW PEOPLE RESPOND WHEN WITNESSING AGGRESSIVE INTERACTIONS

Researcher Name	Ophelia Phillips
Work Address	Department of Clinical, Educational and Health Psychology, University College London, 1-19 Torrington Place, 4th Floor, London WC1E 7HB
Contact Details	email: ophelia.phillips.09@ucl.ac.uk

INVITATION TO TAKE PART

We would like to invite you to take part in our research study. Before you decide we would like you to understand why the research is being done and what it would involve for you. **One of our team will go through the information sheet with you and answer any questions you have.** We'd suggest this should take about 10 minutes. (Part 1 tells you the purpose of this study and what will happen to you if you take part. Part 2 gives you more detailed information about the conduct of the study). Ask us if there is anything that is not clear.

PART 1: DETAILS OF STUDY

Purpose: This research project is looking at people's reactions to virtual environments. This takes the form of cutting edge technology, where a room is filled with 3-Dimensional projections that simulate a real life scenario with moving interactive characters. The aim of the study is to develop our understanding of how people respond when they witness aggressive interactions between other people. We are interested in how people respond physiologically, emotionally and behaviourally. This project is part of a student research project. Please take time to read the following information carefully and ask us if there is anything unclear or if you would like more information. If you are happy to take part in the study you will be asked to sign an informed consent form.

Why have I been invited to take part in this study? You have been invited to take part in this study because we are looking healthy males aged between 18 and 60. We are hoping to recruit twenty participants. The study will take place in the Department of Computer Science, University College London.

Do I have to take part? No, it is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep. You are still free to withdraw at any time, and you do not need to give a reason for this.
What will happen if I decide to take part? Your participation in this project involves five phases and is likely to take between an hour and an hour-and-a-half in total.

Phase 1 - Questionnaires: Before entering the virtual environment, we invite people to complete three questionnaires which ask about your thoughts, feelings and behaviour. There are no right or wrong answers. One of these questionnaires will be given both before and immediately after the virtual environment experience. Completion of questionnaires will take approximately 45 minutes and you will be able to ask for assistance from a member of the research team if you require.

Phase 2 - Physiological measures: We invite people to have their physiological responses measured in two ways; muscle tension and heart rate. To do this we will request to attach two electrodes to you; one to your upper arm and one to your chest. This will be carried out by a male member of the research team who has experience of doing so, in line with standard health and safety conventions. Another member of the research team will also be present during electrode placement for further assistance.

Once this equipment has been attached, we ask participants to enter the virtual reality room and explore the environment that depicts a pub for two minutes. You will be given instructions in the use of virtual reality before entering. Once you are familiar with the use of virtual reality, you will be asked to put on glasses that produce three-dimensional images. During this initial two minute phase, a member of the research team seated outside the virtual reality room will take some measurements of your resting heart-rate and muscle tension using specialist computer software.

Phase 3 - Virtual Reality: After Part 2 has been completed, you will be invited to remain in the virtual pub and continue observing the environment. The scenario will last for two minutes and twenty seconds during which you will be invited to interact with two virtual characters. You will witness an interaction between these two characters; the language used will be realistic to the situation. You will also be asked to respond to the situations and characters to you encounter as you would in everyday life. You will also be asked to wear an Arsenal football team scarf, and to imagine that you are an avid Arsenal supporter.

During your time in the environment a member of the research team will take further muscle tension and heart rate measurements from outside the virtual reality room. **Your behaviour and verbal responses in the virtual environment will be video-recorded** for the purpose of analysing data from the study. Once this has been coded by the researcher, the video tapes will be wiped clean. There will be another researcher directly outside the virtual room at all times to ensure you feel comfortable during the exercise.

Phase 4 – Post-questionnaire and Interview: Once you have completed the exercise you will be asked to complete one of the written questionnaires from Stage 1 a second time. You will then be interviewed briefly about the thoughts and feelings you experienced during virtual reality. Phase 4 is estimated to take 10 minutes. Any data obtained from the interview will be coded. Direct quotes from your interview will NOT be used in the write-up of this study.

Phase 5 – Relaxation and scenario ending: You will be asked to take part in a five minute relaxation exercise, which will be read out to you by one of the researchers. You will then be read a vignette by one of the researchers, explaining how the scenario ended between the two characters in the bar.

Expenses and payment: To thank you for your time, you will be paid £6.00 per 30 minutes of your time spent, which is likely to be 1 hour 30 minutes (totalling £18.00).

What are the possible disadvantages and risks of taking part?
Information about the virtual reality equipment: When people use virtual reality systems they occasionally experience a degree of nausea. If at any time you wish to stop taking part in the study for this or any other reason, please say so and we will stop.

There is some research to suggest that people using virtual reality might experience some disturbance in vision afterwards. No long term studies are known to us, but some studies which have conducted testing after about 30 minutes, have found that the effect is sometimes still there. It is therefore advised that you do not drive a car, motorcycle or operate complicated machinery in the four hours flowing virtual reality. With all kinds of video equipment there is a possibility that an epileptic episode may be generated. **If you have epilepsy please tell us. We would not want you to take part in the study in this case.**

Information about the virtual scenario: It is possible that you may experience a degree of stress or apprehension when witnessing aggression escalating between virtual characters. If at any time you wish to stop taking part for this or any other reason, please tell us and we will stop.

Some researchers have expressed concern that witnessing violence and aggression in virtual reality may increase the likelihood of people becoming aggressive afterwards. To our knowledge no research has directly investigated this. However, recent research into violent video-games and increased aggression in players has found no evidence to support this concern. If however you do notice an increase in aggressive thoughts or urges following your time in the virtual environment and the relaxation exercise, please tell us. In this case we will talk to you about your concerns in order to offer you support.

What are the benefits of taking part? We cannot promise the study will help you but the information we get from this study will help improve the understanding of how people respond to aggressive stimuli. Additionally, some people report that they enjoy taking part in virtual reality, finding it interesting and novel.

What if there is a problem? Any complaint about the way you have been dealt with during the clinical trial or any possible harm you might suffer will be addressed. Detailed information concerning this is given in Part 2 of this information sheet.

Will my taking part in the study be kept confidential? All information from the study will be kept confidential and you will not be identified. As participation is anonymous, it will not be possible for us to withdraw your data once you have returned your questionnaires and taken part in the virtual reality task.

PART 2: ADDITIONAL INFORMATION
What will happen if I don't want to carry on with the study? If you withdraw from the study, we will destroy all your identifiable information, but we may use non-identifiable data that we have collected up until your withdrawal.

What if there is problem? Every care will be taken in the course of this study. However, in the unlikely event that you are injured by taking part, compensation may be available. If you suspect that the injury is the result of the Sponsor's (University College London) negligence then the study is insured by University College London. After discussing with the researcher, please contact Dr Stephen Butler (telephone number: 0207 679 5985), who is the Chief Investigator for the research and is based at the Department of Clinical, Educational & Health Psychology, University College

London.

Will my taking part in this study be kept confidential? All information which is collected about you during the course of the research will be kept strictly confidential and stored in a locked facility.

What will happen to the results of the research study? The results of the research will be analysed as part of a doctorate in clinical psychology and the findings are likely to be published in a scientific journal. Participants will not be identified in any report or publication. Please inform Ophelia Phillips if you would like to be sent a copy of the study's findings.

Who is organising and funding this research? University College London are sponsoring this research.

Who has reviewed this study? All research in the NHS is looked at by independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by UCL Research Ethics Committee.

All data will be collected and stored in accordance with the Data Protection Act 1998.

Consent Form for Clinical Participants:

Version 4: 20.07.2011

PROJECT APPROVED BY NRES COMMITTEE LONDON – CAMBERWELL ST GILES

(To appear on Trust headed paper)

Centre Number:
Study Number:
Patient Identification Number for this trial:
Date:

CONSENT FORM FOR CLINICAL PARTICIPANTS

PROJECT TITLE: A VIRTUAL REALITY STUDY OF HOW PEOPLE RESPOND WHEN WITNESSING AGGRESSIVE INTERACTIONS

Name of Researcher: Ophelia Phillips

	Please initial box
I confirm that I have read and understand the information sheet dated 14.09.2011 (Version 5) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	☐
I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected.	☐
If I decide to withdraw from the study, any identifiable data collected up to this point will be destroyed but non-identifiable data may be used for the research.	☐
I understand that relevant sections of my medical notes and data collected during the study may be looked at this research team at University College London, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.	☐
I understand that I will be interviewed briefly about the thoughts and feelings I experienced during virtual reality. Any data obtained from this interview will be coded and direct quotes from my interview will NOT be used in the write-up of this study.	☐
I agree to my Responsible Clinician being informed of my participation in the study.	☐
I understand that my Responsible Clinician will be contacted if concerns arise about my safety during my participation in the study.	☐

I understand that my verbal responses and behaviour will be video-recorded during my time in virtual reality. These responses will be coded anonymously and will not be identifiable or linked to me in the final report. ☐

I understand that I must not take part if I have epilepsy. ☐

I understand that the information I have submitted will be published as a report and I will be sent a copy if I request this. Confidentiality and anonymity will be maintained and it will not be possible to identify me from any publications. ☐

I understand that I am being paid for my assistance in this research and that some of my personal details will be passed to UCL Finance for administration purposes. ☐

I consent to the processing of my personal information for the purposes of this research study. ☐

I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998. ☐

I agree to take part in the above study. ☐

_____	_____	_____
Name of patient	Date	Signature

_____	_____	_____
Name of person taking consent	Date	Signature

When completed: 1 for participant; 1 for researcher site file; 1 (original) to be kept in medical notes.

Consent Form for Healthy Volunteer Participants

Version 4: 20.07.2011
PROJECT APPROVED BY NRES COMMITTEE LONDON – CAMBERWELL ST GILES AND UCL RESEARCH ETHICS COMMITTEE

Centre Number:
Study Number:
Participant Identification Number for this trial:
Date:

CONSENT FORM FOR HEALTHY VOLUNTEER PARTICIPANTS

PROJECT TITLE: A VIRTUAL REALITY STUDY OF HOW PEOPLE RESPOND WHEN WITNESSING AGGRESSIVE INTERACTIONS

Name of Researcher: Ophelia Phillips

Please initial box

I confirm that I have read and understand the information sheet dated 14.09.11 (version 5) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. ☐

If I decide to withdraw from the study, any identifiable data collected up to this point will be destroyed but non-identifiable data may be used for the research. ☐

I understand that my verbal responses and behaviour will be video-recorded during my time in virtual reality. These responses will be coded anonymously and will not be identifiable or linked to me in the final report. ☐

I understand that I will be interviewed briefly about the thoughts and feelings I experienced during virtual reality. Any data obtained from this interview will be coded and direct quotes from my interview will NOT be used in the write-up of this study. ☐

I understand that I must not take part if I have epilepsy. ☐

I understand that the information I have submitted will be published as a report and I will be sent a copy if I request this. Confidentiality and anonymity will be maintained and it will not be possible to identify me from any publications. ☐

I understand that I am being paid for my assistance in this research and that some of my personal details will be passed to UCL Finance for administration purposes. ☐

I consent to the processing of my personal information for the purposes of this research study. ☐

I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998. ☐

I agree to take part in the above study. ☐

Name of participant	Date	Signature

Name of person taking consent	Date	Signature

When completed: 1 for participant and 1 for researcher site file

APPENDIX 5: POST VR TOOLS

Post VR semi-structured interview schedule

Version 1: 16.07.2011

(To appear on Trust headed paper)

PROJECT APPROVED BY NRES COMMITTEE LONDON – CAMBERWELL ST GILES AND UCL RESEARCH ETHICS COMMITTEE

Post Virtual Reality Semi-Structured Interview

1. Can you briefly describe what you think was happening while you were in the room?

2. How realistic were the unfolding events in the experience that you have just had? Can you rate this on a scale from 1-10, with one being quite unrealistic and easily seen as fantasy, and 10 being quite like real life

3. (a) How did you feel when the first character (the victim) began talking to you?
(b) What did you think of him: did you (a) like, (b) dislike or (c) feel neutral towards him?

4. (a) What did you feel when the second character (aggressor) entered and began speaking to the first? (prompt: e.g. angry/fearful/embarrassed/excited/disgusted etc.)
(b) What was going through your mind?
(c) What did you feel about the two characters at this point (e.g. did you sympathise/feel more affiliated with either?)

5. (a) At which point in the scenario did you experience your strongest emotions
(b) Which emotion?

6. (a) What was your most prominent/overriding emotion during the scenario?
(b) What were the thoughts that were making you feel ……. (repeat the emotion they specified in 6a.

7. Did you feel at any time you wanted to do something about what was happening/intervene?

8. (If they wanted to) What was your main motivation for wanting to intervene during the confrontation (e.g. to intimidate/pacify/protect victim etc.)?

9. (If they did not want to intervene) What would have made it more likely that you would have personally intervened (even though you knew it was virtual reality)?

10. To what extent was your reaction authentic/the same as real life (scale 1-10)?

(To appear on Trust headed paper)

PROJECT APPROVED BY NRES COMMITTEE LONDON – CAMBERWELL ST GILES AND UCL RESEARCH ETHICS COMMITTEE

Progressive muscular relaxation script for participants post-virtual reality

Begin by finding a comfortable position sitting down. You can change positions any time during this exercise to make yourself more comfortable as needed.

Please turn your attention to your breathing. Breathe in forcefully and deeply, and hold this breath. Hold it...hold it... and now release. Let all the air go out slowly, and release all the tension.

Take another deep breath in. Hold it.... and then exhale slowly, allowing the tension to leave your body with the air.

Now breathe even more slowly and gently... breathe in....hold....out....

..breathe in...hold...out...

Continue to breathe slowly and gently. Allow your breathing to relax you.

The next exercise focuses on relaxing the muscles of your body.

Start with the large muscles of your legs. Tighten all the muscles of your legs. Tense the muscles further. Hold onto this tension. Feel how tight and tensed the muscles in your legs are right now. Squeeze the muscles harder, tighter... Continue to hold this tension. Feel the muscles wanting to give up this tension. Hold it for a few moments more.... and now relax. Let all the tension go. Feel the muscles in your legs going limp, loose, and relaxed. Notice how relaxed the muscles feel now. Feel the difference between tension and relaxation. Enjoy the pleasant feeling of relaxation in your legs.

Now focus on the muscles in your arms. Tighten your shoulders, upper arms, lower arms, and hands. Squeeze your hands into tight fists. Tense the muscles in your arms and hands as tightly as you can. Squeeze harder.... harder..... hold the tension in your arms, shoulders, and hands. Feel the tension in these muscles. Hold it for a few moments more.... and now release. Let the muscles of your shoulders, arms, and hands relax and go limp. Feel the relaxation as your shoulders lower into a comfortable position and your hands relax at your sides. Allow the muscles in your arms to relax completely.

Focus again on your breathing. Slow, even, regular breaths. Breathe in relaxation.... and breathe out tension..... in relaxation....and out tension.... Continue to breathe slowly and rhythmically.

Now focus on the muscles of your buttocks. Tighten these muscles as much as you can. Hold this tension..... and then release. Relax your muscles.

Tighten the muscles of your back now. Feel your back tightening, pulling your shoulders back and tensing the muscles along your spine. Arch your back slightly as you tighten these muscles. Hold.... and relax. Let all the tension go. Feel your back comfortably relaxing into a good and healthy posture.

Turn your attention now to the muscles of your chest and stomach. Tighten and tense these muscles. Tighten them further...hold this tension.... and release. Relax the muscles of your trunk.

Finally, tighten the muscles of your face. Scrunch your eyes shut tightly, wrinkle your nose, and tighten your cheeks and chin. Hold this tension in your face.... and relax. Release all the tension. Feel how relaxed your face is.

Notice all of the muscles in your body.... notice how relaxed your muscles fee l. Allow any last bits of tension to drain away. Enjoy the relaxation you are experiencing. Notice your calm breathing.... your relaxed muscles.... Enjoy the relaxation for a few moments....

When you are ready to return to your usual level of alertness and awareness, slowly begin to re-awaken your body. Wiggle your toes and fingers. Swing your arms gently. Shrug your shoulders. Stretch if you like.

You may now end this progressive muscle relaxation exercise feeling calm and refreshed.

Vignette – shaking hands and making up
Version 1: 14.09.2011

PROJECT APPROVED BY NRES COMMITTEE LONDON – CAMBERWELL ST GILES

(To appear on Trust headed paper)

After you left the two characters in the pub, a third, neutral character entered who was not affiliated with any particular football team. This character intervened between the two men and asked them what had been happening. The three of them sat at a table and talked calmly about what had happened. The person who initially started the confrontation apologised to the other character for being rude. The apology was accepted and both men shook hands, leaving the bar with no hard feelings. Do you have any questions about this?

APPENDIX 6: BEHAVIOUR CODING CHART TEMPLATE

Participant ID	Number of conciliatory/ prosocial interventions		Number of antagonistic/ aggressive interventions	
	Non-verbal	Verbal	Non-verbal	Verbal