Making an IMPACT: Designing and Testing a Novel Attentional Training Game to Reduce Social Anxiety

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters
Making an IMPACT: Designing and Testing a Novel Attentional Training Game to Reduce Social Anxiety

A dissertation presented

by

Philip Miles Enock

to

The Department of Psychology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Psychology

Harvard University

Cambridge, Massachusetts

April 2015
Abstract

Development of novel candidate interventions to treat anxiety disorders is an important research priority, given the burden of these disorders, barriers to treatment access, and the promising but limited success of current approaches, including attentional bias modification treatment. I created a novel training game paradigm, Intrinsically-Motivating Playable Attentional Control Training (IMPACT), with several potential ways that its design could increase the strength of attentional change and commensurate clinical benefits beyond existing training methods.

In a large online experiment, I randomized participants among three alternative IMPACT training conditions. All involved the same smiling and disgust faces falling down on the screen, and players tapped faces to score points and prevent them from reaching the bottom. In IMPACT-Positive, players tapped smiling faces only, ignoring disgust faces. In IMPACT-Threat, players tapped disgust faces, ignoring smiling faces. In IMPACT-Undirected, players tapped all faces without regard to expression. After training, participants completed flanker tasks, reaction-time measures of general and emotional attentional control and attentional bias toward threat versus neutral stimuli. Participants also confronted an anxiety-provoking stressor and rated their state anxiety before IMPACT, after IMPACT, and after the stressor.

I tested hypotheses regarding differential effects of the training variants on attentional measures and anxiety reactivity, finding that training did not cause group differences in measures of general or emotional attentional control, but they did lead to differences in attentional bias. The anxiety-provoking stressor induced a rapid rise in anxiety, but no differences emerged among the training conditions.
Overall, results show the potential for researchers to abandon the tradition of repeated reaction-time trials in favor of engaging, fluid games that continuously motivate trainees and prompt attentional shifts. Additional testing of the IMPACT paradigm is needed to establish whether this particular game training approach is clinically useful for reducing anxiety.
# Table of Contents

<table>
<thead>
<tr>
<th>Abstract</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>vii</td>
</tr>
<tr>
<td>Chapter 1: Background</td>
<td>1</td>
</tr>
<tr>
<td>Computerized treatments for anxiety</td>
<td>1</td>
</tr>
<tr>
<td>Research priorities for computerized treatment research</td>
<td>1</td>
</tr>
<tr>
<td>Limited-capacity attentional processing</td>
<td>3</td>
</tr>
<tr>
<td>Attentional bias toward threat in anxious individuals</td>
<td>3</td>
</tr>
<tr>
<td>Attentional control deficits in anxious individuals</td>
<td>5</td>
</tr>
<tr>
<td>Literature review of attention bias modification, a candidate treatment for reducing social anxiety</td>
<td>6</td>
</tr>
<tr>
<td>Chapter 2: Design considerations for IMPACT, a novel attentional training game</td>
<td>8</td>
</tr>
<tr>
<td>Description of IMPACT active training</td>
<td>9</td>
</tr>
<tr>
<td>Sources of design inspiration: Games, not tasks</td>
<td>10</td>
</tr>
<tr>
<td>Potential attentional bias effect: Training a bias toward tapped stimuli</td>
<td>11</td>
</tr>
<tr>
<td>Potential attentional control effects: General or valence-specific attentional control</td>
<td>11</td>
</tr>
<tr>
<td>More frequent attentional shifts</td>
<td>11</td>
</tr>
<tr>
<td>Use of wider range of stimuli</td>
<td>12</td>
</tr>
<tr>
<td>Three axes of motivation enhancement</td>
<td>12</td>
</tr>
<tr>
<td>Chapter 3: Effect of a novel attentional training game on anxiety reactivity to a stressor in a single-session experiment</td>
<td>14</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------</td>
<td>------</td>
</tr>
<tr>
<td>Introduction</td>
<td>14</td>
</tr>
<tr>
<td>Method</td>
<td>15</td>
</tr>
<tr>
<td>Materials</td>
<td>17</td>
</tr>
<tr>
<td>Procedure</td>
<td>21</td>
</tr>
<tr>
<td>Data Analyses</td>
<td>22</td>
</tr>
<tr>
<td>Results</td>
<td>24</td>
</tr>
<tr>
<td>Discussion</td>
<td>33</td>
</tr>
<tr>
<td>References</td>
<td>41</td>
</tr>
<tr>
<td>Appendix A: Speech anticipation stressor</td>
<td>48</td>
</tr>
</tbody>
</table>
Acknowledgments

In this section, I thank a subset of people helpful to my research and graduate school life. The list is not exhaustive. Also, the importance of these individuals’ contributions transcends what I can convey in a stiff academic document whose production is required for graduation.

For guidance and mentoring, I thank my advisor, Richard McNally.

For collaboration on IMPACT (this dissertation project), I thank my collaborators (who will be co-authors for an eventual journal article) Donald Robinaugh, Ernst Koster, Rudi De Raedt, Howard Braham, and Richard McNally.

For contributions to my first major research project (Enock, Hofmann, & McNally, 2014), I thank Jennifer Yu, Imke Vonk, Linh Vu, and Joshua Breckman.

For service on my dissertation committee, I thank Richard McNally, Matthew Nock, Joshua Buckholtz, and Stefan Hofmann.

For helping me prepare for graduate school, I thank Ethan Kross, Regina Miranda, and Walter Mischel.

For friendship and scientific feedback, I thank my labmates Donald Robinaugh, Dianne Hezel, Nicole Leblanc, Emily Bernstein, Hannah Reese, and CA Meyenburg, as well as my colleagues elsewhere, Michelle Rozenman, Alexandre Heeren, and Peter Clasen.

For hosting my visits and bringing me into an international research community, in addition to scientific discussions, I thank Colin MacLeod and his lab members, as well as Per Carlbring, Reinout Wiers, Rudi De Raedt, Ernst Koster, Marcella Woud, Anne-Wil Kruijt, and Eva de Hullu.

For statistical instruction, I thank Samuel Moulton, Matthew Killingsworth, James Sidanius, Arnold Ho, Delphine Courvoisier, and Patrick Mair.
For fostering a sense of community in the psychology department, I thank William Santoro, James “Jimmy” Robson, Karim Kassam, Adrian Ward, Anna Leshinskaya, and everyone who hosted parties. Also for relevant communities, I thank all Dudley House fellows and organizers, as well as Timothy Dransfield, Anthony Corr, Andrew Greenspon, Paul Hess, and especially the distinguished and indomitable William Steinhardt.

For tireless and effective administration of the graduate program in psychology, I thank Celia Raia.

For friendship and support, I thank colleagues of Party Cohort #1, a moniker that represents my wisest aspiration in graduate school, as ambitious as my research efforts. Its membership will entirely, henceforth, regardless of individuals’ participation levels, be enshrined in Harvard’s hallowed archives and ProQuest’s dusty servers in perpetuity as Anna Leshinskaya, Donal Cahill, Jennifer Sheehy-Skeffington, Juan Manuel Contreras, Roman Feiman, Jordan Suchow, Steven Frankland, Christine Cha, Mei Yi Ng, Nancy Lau, Kathryn Hobbs, Su Keun Jeong, Jon Chamberlain, Amanda Ie, and Michael Cohen. I thank all of my other graduate student colleagues, additionally.

I thank all of my friends elsewhere, and all of my family.

I thank Katie Porter for love and for being my everything.
Chapter 1: Background

Computerized treatments for anxiety

Anxiety disorders constitute a major burden on public health and are the most prevalent type of mental disorder (Kessler et al., 2007). Current treatment approaches, such as psychotherapy and psychopharmacology, are somewhat effective, yet access to and utilization of these treatments are limited (Kessler et al., 2010). Nationwide, there is a severe lack of access to psychotherapists who perform the most effective forms of psychological treatment for anxiety disorders, and expanding the use of face-to-face talk therapy seems inadequate for addressing the unmet need for treatment (Kazdin & Blase, 2011; Kazdin & Rabbitt, 2013).

Computerized treatment methods enable greater accessibility and lower cost than face-to-face psychotherapy. Computerized cognitive behavior therapy (CBT) has reduced anxiety symptoms with large effect sizes (Andrews, Cuijpers, Craske, McEvoy, & Titov, 2010), and these approaches are routinely used in the UK’s national health system to treat certain anxiety disorders and depression (National Institute for Health and Clinical Excellence, 2010), although their use is less common in the US.

Research priorities for computerized treatment research

It is useful to consider what directions and processes of research will lead to improved treatments and the reduction of the global burden of mental health over the long run (for further discussion of the treatment research process, see Enock & McNally, 2013). Three crucial research directions are: (1) rigorously evaluating existing candidate interventions, (2) expanding accessibility to efficacious interventions through new technology such as mobile devices, and (3) developing and testing novel candidate treatments.
For existing computerized interventions that may be efficacious, randomized controlled trials (RCTs) are needed to compare the active intervention to a control condition to establish efficacy for symptom reduction. RCTs of psychological interventions have tended to be small, with sample sizes of 20–40 per condition. When the true effect sizes of symptom reductions from an active treatment versus control are small to medium, as is so often the case, undersized samples can exacerbate the problem of non-replications with negative findings (Type II error) due to low statistical power, making it difficult to identify what variability in outcomes is due to methodological differences among studies or due to sampling error. Thus, larger RCTs testing computerized interventions are needed.

Technology may expand access to computerized interventions. Since the first iPhone was introduced in 2007, smartphones have continued to grow in popularity and capabilities. The always-on-hand nature of smartphones makes them well-suited to treatment delivery when frequent use is clinically desirable. People have shown a willingness to use mobile apps (whether on smartphones or tablets) for diverse purposes such as productivity, information lookup, and games, with the number of global mobile app downloads having reached 18 billion in 2011 (Portio Research, 2012). Given this frequent user experimentation, people may be more apt to use psychological interventions via mobile app than via personal computer.

Development of novel candidate treatments is another important research direction. While existing candidate treatments should be evaluated thoroughly, novel approaches must also be generated to find opportunities for greater efficacy. Furthermore, if novel treatments have alternative mechanisms of action, this allows for augmentation of current approaches with additive benefits. High-risk research and development, where novel methods are less assured of efficacy but open up new possibilities if efficacious, is less frequently pursued than incremental research, and it is much-needed.

Aside from the clinical goal of increasing efficacy of available treatments, novel methods also enable novel insights into mechanisms involved in the etiology and maintenance of anxiety disorders. The
mechanisms of attentional bias (AB) and attentional control (AC) are potential intervention targets, and manipulating these mechanisms in experiments also allows for probing their causal impact on anxiety.

**Limited-capacity attentional processing**

Attentional processing enables human beings to receive massive amounts of sensory input and make use of relevant aspects of this information. Since processing all input to the fullest extent possible is unfeasible (Broadbent, 1958), attention involves selective processing of information and limited-capacity mechanisms of information processing, as observed in the well-studied system of visual attention (e.g., Desimone & Duncan, 1995). The attentional blink paradigm demonstrates a limited capacity scenario in temporal attention: In a rapid serial visual presentation (RSVP) task, where subjects view a series of letters one-by-one and must report when digits (targets) appear in this stream, subjects show a diminished ability to report a second target if it appears shortly (200-500 ms) following a first target (Dux & Marois, 2009). A potential explanation of this phenomenon is that mechanisms of visual attention are capacity-limited. Many different theoretical accounts for this phenomenon are available (Marois & Ivanoff, 2005), but a simplified explanation is that attending to the first target leaves less attentional capacity available for attending to the second.

**Attentional bias toward threat in anxious individuals**

Researchers have identified information-processing anomalies in people with severe, chronic anxiety, including those with diagnosable anxiety disorders. Anxious individuals show AB for threatening stimuli (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & van IJzendoorn, 2007), whereby they attend preferentially to even mildly threatening items over neutral or positive stimuli. AB is exhibited in automatic, stimulus-driven (bottom-up) attentional processing, such that threatening stimuli are more salient than other stimuli. In everyday life, it is crucial for an individual to attend urgently to a speeding car while crossing the street. This would be a case of stimulus-driven attention appropriately trumping whatever other goals the individual was pursuing at the time. The most adaptive deployment of attention
is toward the threat to survival. However, frequently paying undue attention to more minor threatening stimuli instead of neutral or positive stimuli may unnecessarily increase individuals’ propensity to experience high anxiety. Notably, the limited-capacity nature of attention implies that individuals attending to threatening stimuli are less able to attend to other stimuli.

Computerized reaction-time tasks such as the dot-probe paradigm (MacLeod, Mathews, & Tata, 1986), emotional Stroop (Mathews & MacLeod, 1985), and spatial cueing (Fox, Russo, Bowles, & Dutton, 2001) tasks have documented AB in groups of individuals with high trait anxiety compared to groups with low trait anxiety. For example, in the original dot-probe task, participants perform repeated trials, each time viewing a pair of words (one neutral, one threatening), one above and one below center screen. The words then disappear, and, on most trials, the screen is blank for 1 s, followed by the next word pair. On 25% of the trials, however, a small dot replaces one of the words. Upon seeing this dot probe, participants press a button as quickly as possible, signifying detection of the dot. People with high trait anxiety are often faster to respond to the dot when it replaces a threat word than when it replaces a neutral word, relative to people with low trait anxiety. Hence, an attentional bias for threat is inferred from rapid detection of a dot in a location vacated by a threat cue. The probe-discrimination task is a variant of the dot-probe task. Instead of pushing a single button to indicate detection of a dot, participants push one of two buttons to indicate the identity of a probe (e.g., E or F). People with high trait anxiety are often faster to discriminate probes when they replace threat cues than when they replace neutral cues, relative to people with low trait anxiety.

MacLeod (1995) also realized that the dot-probe task could be altered to induce a bias. If probes repeatedly followed the location of neutral stimuli, opposite the threat location, then performing this task might induce a bias away from threat. Clinically, inducing a bias away from threat could potentially reduce chronic anxiety and be applied as a treatment for anxiety disorders. This approach of attentional bias modification (ABM; MacLeod, Rutherford, Campbell, Ebsworthy, & Holker, 2002) has become a much-studied existing candidate intervention for anxiety.
Attentional control deficits in anxious individuals

Goal-directed (top-down) and stimulus-driven (bottom-up) attentional processes compete with each other. In an anxiety-relevant example, a socially anxious person attempting to engage in a conversation is maintaining the goal of interacting with the conversation partner. This top-down attention, also called cognitive control or executive control, involves the prefrontal cortex’s propagating signals to other areas of the brain to influence behavior in service of the goal (Miller & Cohen, 2001). Simultaneously, stimulus-driven attention may be captured by external threat cues or internal ones (such as worries), distracting from the goal-directed processes of the social interaction. Thus, an individual preoccupied with being negatively evaluated by others may have difficulty disengaging attention from threat, making it difficult to attend to the ongoing conversation as desired. Experimental evidence documents this difficulty disengaging from social threat cues in socially anxious individuals (Buckner, Maner, & Schmidt, 2010). Greater AC, involving a facility with voluntarily shifting attention, would aid the person in shifting attention away from concerns and towards a sustained focus on the conversation. AC would aid in the most adaptive deployment of attention, which, in this social context, would be to favor the goal-directed attentional system and attend less to the distracting social concerns.

In an AC theory account (Eysenck, Derakshan, Santos, & Calvo, 2007), high trait anxiety is associated with a general deficit in AC, exemplified by diminished ability to ignore irrelevant stimuli that distract from a primary cognitive task. By this account, heightened anxiety renders the goal-directed attentional system weaker while increasing the influence of the stimulus-driven attentional system. The AB and AC accounts are not necessarily in conflict, but rather represent interrelated lenses through which to relate attentional phenomena to anxiety. Indeed, greater AC may effectively suppress AB toward threat (Derryberry & Reed, 2002), which is a stimulus-driven phenomenon.

Multiple interventions have emerged targeting the AC pathway, such as meta-cognitive therapy for depression (Wells et al., 2012), mindfulness training for depressive relapse prevention (Teasdale,
Segal, & Williams, 1995), and cognitive control training (Siegle, Ghinassi, & Thase, 2007). Given that these approaches have gained traction in the domain of depression treatment, combined with theoretical background of AC in anxiety, exploring computerized interventions to target AC is a promising research direction.

Although ABM was conceived to modify AB, it may well alter AC instead or in addition to diminishing AB (Heeren, De Raedt, Koster, & Philippot, 2013). To maximize clinical efficacy, it may be desirable to target both pathways in a candidate intervention. By taking both approaches into account when designing a novel intervention, it may be possible to more strongly affect both pathways than do current training methods conceived to influence only one or the other.

**Literature review of attention bias modification, a candidate treatment for reducing social anxiety**

For ABM, the prevailing training methods have always been variants of the dot-probe task (MacLeod et al., 2002). In the popular probe-discrimination variant, a trial consists of a fixation cross appearing at center screen for 500 ms, followed by a pair of stimuli (either words or faces), one threatening and one neutral, appearing for 500 ms. Then, the stimuli disappear, and one of them is replaced by a small probe, such as one or two dots, or an E or F. Participants must identify the probe by pressing a corresponding key, thus inducing their attention to the recent location of one of the stimuli. In the active training condition, probes always replace the neutral stimuli, whereas in control training, 50% of probes replace the neutral stimulus and 50% replace the threat stimulus. The clinical domain attracting the largest number of ABM treatment studies has been social anxiety, and the stimulus pairs have been a pair of faces, one displaying a neutral expression and the other displaying disgust. Disgust faces convey contempt and are rated as threatening by socially anxious people (Amir, Najmi, Bomyea, & Burns, 2010).

In the first two RCTs in people diagnosed with generalized social anxiety disorder (Amir et al., 2009; Schmidt, Richey, Buckner, & Timpano, 2009), participants were randomly assigned to receive eight sessions of ABM or control training, which occurred in the laboratory twice per week for four
weeks. These studies showed very strong benefits for ABM over control training: 50% or more of ABM participants no longer met disorder criteria by the end of training, whereas less than 20% of control participants no longer qualified for the diagnosis.

In the context of these promising results, I sought to augment the seemingly-highly efficacious ABM treatment via increased dosage. I also wished to make it more palatable to users, and to test whether training would function in a smaller format and in locations outside of the laboratory. To achieve these goals, I designed the first study to employ smartphone-delivered ABM training (Enock & McNally, 2010, as summarized in Enock & McNally, 2013). The key goal was to deliver more frequent training, amounting to a higher dose, defined as a greater number of training trials during the course of treatment. More frequent, brief learning sessions are more suited to skills training than are infrequent, longer sessions (Bjork & Bjork, 2011). Having briefer sessions could also reduce participant boredom, which, from widespread anecdotal reports, had always been a concern for dot-probe training research. In a multiple baseline across subjects experimental design (Barlow, Nock, & Hersen, 2009), 16 participants trained thrice daily on their smartphones, performing 1 or 2 weeks of control training followed by 3 weeks of ABM training. Anxiety self-report scales revealed significant declines during the protocol, but the degree of declines were not significantly different during the control versus ABM training periods, thus failing to support the efficacy of ABM over control training for reducing anxiety.

Since the previous study had a small sample and non-significant results, I conducted an RCT with a larger, Internet-recruited sample of 326 participants (Enock, Hofmann, & McNally, 2014). Participants were randomly assigned to perform four weeks of ABM or control training, and a smaller proportion was allotted to the waitlist group. Both ABM and control training groups showed significantly greater declines in social anxiety scores than did waitlist, showing that training confers clinical benefits. However, benefits were statistically indistinguishable between the active and control groups. In terms of AB, measured via dot-probe task weekly on smartphones, ABM induced a significantly greater bias away from threat compared with control, though the effect was small.
Other RCTs of ABM versus control training for reducing social anxiety have appeared in recent years and shown similar results, in that they have tended to show no differences in symptom reduction for ABM versus control training (Boettcher et al., 2013; Boettcher, Berger, & Renneberg, 2012; Bunnell, Beidel, & Mesa, 2013; Carlbring et al., 2012; McNally, Enock, Tsai, & Tousian, 2013; Neubauer et al., 2013; Sawyer et al., 2012), although one study did show ABM to be superior to control (Heeren, Reese, McNally, & Philippot, 2012).

Taken together, ABM research suggests that dot-probe ABM training can confer symptom reduction, yet the effect is usually small (g = 0.27 for reducing social anxiety, according to a meta-analysis, Heeren, Mogoaşê, Philippot, & McNally, under review), and its efficacy may be contingent upon certain implementation parameters that have not been identified. Especially worrisome is that effect sizes appear smaller in recent versus early studies (publication year significantly moderated effect size in Heeren et al., under review). Given the limited strength of dot-probe ABM training, incremental optimizations would be inadequate; thus, I chose to create a novel game paradigm, representing a sea change in training approach with several potential axes of improvement over existing methods.

Chapter 2: Design considerations for IMPACT, a novel attentional training game

I created a novel training paradigm, Intrinsically-Motivating Playable Attentional Control Training (IMPACT), with several potential ways that its design could help to outperform the dot-probe for attentional change and clinical impact. It is only with these strong rationales for improvements that the great cost in time and funds involved in creating the game would be warranted. IMPACT is not a gamified version of ABM; rather, it is a fundamentally different approach, albeit also an attentional training method.

I will describe the IMPACT paradigm and discuss its potential advantages. For simplicity, I will maintain the context of the active training condition, where players tap (or click) smiling faces while
ignoring disgust faces. In actual experiments using IMPACT, there will be multiple training variants, such as tapping disgust faces while ignoring smiling ones, or tapping all faces.

**Description of IMPACT active training**

In IMPACT, smiling and disgust (i.e., social threat) faces continuously descend from the top to the bottom of the screen. The following describes the active version of IMPACT training, whereas other variants differ in limited ways. Smiling faces constitute 60% of the faces that appear, while disgust faces constitute the remaining 40%. The essence of the active training condition is that players must click or tap smiling faces to score points. For gameplay on smartphone or tablet devices, the action is tapping with a finger or thumb, whereas players using computers and mice or trackpads click the left mouse button. Only smiling faces are relevant to the game and point-scoring, making them worthy of players’ attention while disgust faces function solely as distractors. That is, the game incentivizes players to ignore disgust faces.

Once the player taps a smiling face for the first time, the face bounces upward a short distance, during which time it is untappable for about 0.5-1 s. Then, the face becomes tappable again. Once the player taps that face a second time, the face bounces all the way off the top of the screen, disappearing. Essentially, the goal of the game is to prevent smiling faces from reaching the bottom of the screen by tapping them. Players receive +1 point for tapping a smiling face the first time, and then they receive +5 for tapping the same face a second time after it is once again tappable. Players lose 3 points for every smiling face that falls off the bottom of the screen.

Disgust faces have no effect on score. They simply continue to fall regardless of the player’s actions. If a player taps a disgust face, a black “X” appears for a short duration to signal to the player that this face cannot be tapped. The same black “X” appears if the player taps a smiling face while it is temporarily untappable.
To prevent players from employing a strategy of tapping all faces regardless of valence, the game enforces a rate limit. If a player taps a disgust face or empty space more than three times within 4 s, the screen turns red for 1.5 s and blocks all input. Thus, players learn to tap accurately to avoid incurring a penalty from having a face fall off the bottom of the screen.

**Sources of design inspiration: Games, not tasks**

The emotional dot-probe task (MacLeod et al., 1986) was an adaptation of a visual attention task (Navon & Margalit, 1983), in the tradition of vision research (e.g., Posner, 1980). These tasks were designed for the purpose of identifying locations on screen that were more likely capturing participants’ attention, without regard for inspiring much motivation. Many participants find dot-probe training to be “repetitive and boring” (Beard, Weisberg, & Primack, 2012).

For IMPACT, the design is inspired by games rather than attentional measurement tasks. The important step was recognizing that there are multitudes of ways to induce participants to shift their attention. The constraint of the dot-probe’s design was that it needed to measure attention. As such, it is a trial-based task, with stimulus presentation repeated many times with the same central fixation cross, a stimulus pair in the same screen locations at each trial, and two potential probe locations tied to those stimuli. Novel training tasks (e.g., Clarke, MacLeod, & Guastella, 2011) have tended to continue in the tradition of repeated trial-based tasks, in order to be able to fulfill the requirements of both training and measurement. In contrast, the constraint disappears when designing a task for training only and not measurement. For a training task, there is no need to adhere to repeated trials with stimuli in predictable locations. Separating the training requirement from the measurement requirement allowed games, in contrast to measurement tasks, to serve as the basis for design. The games Reflex (Team Missionred, 2001) and a genre of games known as juggling games inspired IMPACT.
Potential attentional bias effect: Training a bias toward tapped stimuli

By training players to tap smiling faces, the active condition may induce an AB towards positive stimuli. Players must seek positive stimuli and directly interact with them by tapping (or clicking), rendering the game more directly engaging than the dot-probe task whereby participants attend to a neutral probe that merely replaces a positive (or neutral) stimulus that is no longer present on the screen.

Potential attentional control effects: General or valence-specific attentional control

IMPACT training could increase players’ ability to ignore task-irrelevant stimuli in general, constituting a general AC effect. Yet effects may well be more specific: Since players are practicing ignoring one valenced class of emotional stimuli and attending to another, the effects may be valence-specific. Emotional AC refers to AC involving valenced stimuli specifically. In order to play effectively and score more points, players need to avoid wasting time looking at threat faces while searching for positive ones. The most efficient attentional deployment pattern incentivized by the game is to attend only to positive stimuli, ignoring threat faces. Thus, the mechanics of the game drive players to learn not to attend (and to disengage from) threat stimuli. The active training game incentivizes the inhibition of attention to threat stimuli within the game. Hopefully, such a training generalizes to affect threat processing outside of the game (emotional AC), or to participants’ ability to focus on goal-relevant stimuli while ignoring distractors regardless of whether threat is involved (general AC).

More frequent attentional shifts

Attentional training paradigms aim to provide participants with practice in attentional shifting. In a standard dot-probe ABM procedure, each trial lasts for over 2 s, hence the training repetition for attentional shifts occurs only once every 2 s. In a fluid, continuous training paradigm, such as IMPACT, attentional shifts are much more frequent. Multiple times within even one second, the trainee needs to shift attention towards smiling faces and shift away from threat faces. This creates a potential avenue for
increased training efficacy in IMPACT versus dot-probe paradigms, even when training session duration is equivalent. Efficacy should rest on the number of shifts, not task duration.

Use of wider range of stimuli

ABM research has typically employed a small number of face stimuli (e.g., 8 stimulus pairs in Amir et al., 2009; Carlbring et al., 2012; Schmidt et al., 2009). However, when aiming to train on a limited set of stimuli and generalize the training effects to all threat stimuli encountered by trainees, a much larger set is warranted. Furthermore, stimuli within one experimental stimulus set are uniform in many ways, such as lighting, exact size, and sometimes ethnicity. Training with more diverse stimuli makes sense, so that effects are not limited to stimuli with certain characteristics. Accordingly, I incorporated 143 face models into IMPACT by employing three different stimulus sets.

Three axes of motivation enhancement

Motivation to train for long periods

Since high doses of training may be clinically beneficial, a training paradigm should be intrinsically motivating, increasing its users’ willingness to use it for long periods. Designing a training based on enjoyable games avoids creating a bitter medicine that users find dull (Beard et al., 2012) and instead creates an experience that users wish to repeat.

Motivation by appropriate challenge

Tasks that are challenging are more engaging to users, though only up to the point before the task is discouragingly difficult. IMPACT is designed to adapt to players’ performance between every one-minute round of play: If the player excelled in tapping smiling faces in a given round, then the game provides a faster-paced next round with more faces. By contrast, a player who struggled in the first round receives a slower-paced next round with fewer faces. The game makes larger adjustments in earlier
rounds and smaller ones in later rounds, ensuring that the game quickly arrives at an appropriate difficulty level for each player early on, while providing more stability in later rounds. Thus, the game is challenging for players with extensive general gaming experience (or practice at IMPACT) and is still equally playable for inexperienced players. Adaptive difficulty enables games to reach even populations, such as older adults, who are less likely to play games recreationally. Calibrating task demands ensures a level of “desirable difficulty” that psychologists have found fosters steady, robust learning of diverse skills (Bjork & Bjork, 2011).

**Motivation to perform attentional shifts at each moment**

Training is presumed to occur through repeated attentional shifts by participants. In dot-probe training, participants’ motivation for shifting responding to probes may be weak, stemming primarily from their desire to follow the experimenter’s directions. Successful games inherently create powerful and intrinsic motivation: They can provide reinforcement, in the form of points, at each moment of action. This makes participants more likely to perform the attentional shifts and to harness their cognitive resources towards carrying out these actions to a greater extent than for a less motivating task. Goal-directed behavior is induced by the continuous reinforcement (constantly-available point-scoring) and pursuit of high achievement in the game (a motivating goal for many game-playing individuals), as well as avoidance of negative outcomes (losing points). In comparison, dot-probe training lacks sufficient reinforcement. Upon each trial of dot-probe, the response to identify a probe is not substantially rewarded, and participants may be most motivated only by wishing to finish the experiment to pursue other activities or to follow directions to please the experimenter. Accordingly, IMPACT capitalizes on operant conditioning, in contrast to traditional dot-probe paradigms’ having close ties to neo-Pavlovian associative learning.
Chapter 3: Effect of a novel attentional training game on anxiety reactivity to a stressor in a single-session experiment

Introduction

Single-session experiments are the most effective approach for a candidate intervention at this early stage of development and testing. They offer greater feasibility compared with multi-week treatment trials, and thus provide opportunities to test the paradigm, observe its effects, modify as necessary, and then proceed with further experiments. In an anxiety inoculation experimental design (e.g., MacLeod et al., 2002), participants first receive active or control training to induce differences in attentional performance between the groups. Attentional measurement tasks test whether the attentional manipulation was successful. Then, participants confront an anxiety-provoking stressor. The data show whether active training served as an inoculation to attenuate increased anxiety relative to control.

Following this framework, the present experiment randomized participants among three alternative IMPACT training conditions. All involved the same smiling and disgust faces falling down on the screen, and players tapped faces to score points and prevent them from reaching the bottom. In IMPACT-Positive, players tapped smiling faces only, ignoring disgust faces. In IMPACT-Threat, players tapped disgust faces, ignoring smiling faces. In IMPACT-Undirected, players tapped all faces without regard to expression.

After training, participants completed flanker tasks, reaction-time measures of AC and AB. I relied on post-training only assessment, omitting pre-training assessment in order to reduce participant burden and maintain a concise procedure. A flanker task using arrows indexed general AC, and a flanker task using emotional faces measured emotional AC. I opted not to include a dot-probe measure of AB, due to its near-zero reliability in many reports (Ataya et al., 2012; Enock et al., 2014; McNally et al., 2013; Schmukle, 2005; Staugaard, 2009; Waechter, Nelson, Wright, Hyatt, & Oakman, 2013). I did extract AB scores from the flanker task using faces, representing an alternative to visual probe methods.
After the flanker tasks, participants confronted an anxiety-provoking stressor, namely, the instruction that they would need to give a 3-min speech with only 1 min to prepare. Participants rated their state anxiety before IMPACT, after IMPACT, and after the speech instruction.

The experiment tested the following a priori hypotheses: (1) whether IMPACT-Positive and IMPACT-Threat would increase general AC relative to IMPACT-Undirected, (2) whether IMPACT-Positive would affect emotional AC in facilitating the ignoring of threat relative to IMPACT-Threat and IMPACT-Undirected, (3) whether IMPACT-Threat would make threat easier to attend to relative to IMPACT-Undirected, (4) whether IMPACT-Positive would reduce anxiety reactivity relative to IMPACT-Threat and IMPACT-Undirected, and (5) whether IMPACT-Threat would increase anxiety reactivity relative to IMPACT-Undirected. Hypothesis 6 was whether IMPACT-Threat would induce an AB toward threat stimuli relative to IMPACT-Positive and IMPACT-Undirected, and Hypothesis 7 was whether IMPACT-Positive would induce an AB away from threat stimuli relative to IMPACT-Undirected. Hypotheses 6 and 7 were added only after data collection had begun, when I identified an opportunity to calculate the AB score from AC task data.

Method

Participants

Data were collected between October 6 and October 12, 2014. All participants were recruited via Amazon Mechanical Turk (MTurk; Shapiro, Chandler, & Mueller, 2013) and were compensated $6 each. The study listing encouraged individuals with iOS or Android smartphones or tablets to sign up, though it informed them that any computer with the Google Chrome browser was acceptable. Minimal inclusion criteria were applied: I configured MTurk’s screening controls to ensure that participants were in a United States location and that over 95% of the individual’s previous MTurk jobs completed had been approved by the job requester. Participants had to confirm prior to the consent form that they were fluent in English. As part of consent, participants confirmed that they were at least 18 years of age and did not have repetitive strain injury or carpal tunnel syndrome relating to mouse clicking or touchscreen tapping.
Of those who viewed the posting, 696 individuals agreed to the consent form to begin the study. Of these consenters, 673 continued through the initial questionnaires and were randomized to game conditions. Of those randomized, 621 completed the procedures and submitted compete data. Of these, 569 surpassed an accuracy cutoff of 80% correct trials on both flanker tasks, which I required for inclusion in the completer sample used for all analyses. Hence, the completer sample comprised 569 participants (i.e., 84.5% of those randomized, and with 15.5% dropout rate). Although I attempted to obtain reasons for any dropouts from the study and created a survey and separate URL for participants needing to withdraw, too few participants (n = 8) reported a reason, rendering analysis meaningless. The completer sample was 51.3% male. Race/ethnicities were 81.7% White or Caucasian, 8.3% Black or African-American, 7.2% Hispanic or Latino, 4.7% Asian, 1.9% American Indian or Alaskan Native, 0.2% Native Hawaiian or Pacific Islander, and 0.4% Other. Education, in years starting with the first grade, had M = 15.2 (SD = 2.8). Participants used touchscreen (62%) or non-touchscreen (38%) devices for game training and flanker tasks (for device types, see Table 1). Characterizing the sample on anxiety self-report measures, the Penn State Worry Questionnaire (PSWQ; Meyer, Miller, Metzger, & Borkovec, 1990) showed M = 48.8 (SD = 17.6), and the Social Interaction Anxiety Scale (SIAS; Mattick & Clarke, 1998) showed M = 31.3 (SD = 20.0). For comparison, a previous study’s sample of individuals diagnosed with social anxiety disorder showed M = 56.5 on the SIAS, with its healthy control sample’s showing M = 15.5 (Sposari & Rapee, 2007).

Random assignment led to training group sizes of n = 209 (IMPACT-Positive), n = 172 (IMPACT-Threat), and n = 188 (IMPACT-Undirected). A chi square test confirmed that this distribution of membership did not significantly differ from the expected random assignment of 1/3 per group (χ²(2, N = 569) = 3.63, p = .16).
Table 1. Device types used for IMPACT training and flanker tasks.

<table>
<thead>
<tr>
<th>Device Type</th>
<th>n</th>
<th>Percentage of sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smartphone</td>
<td>229</td>
<td>40.2%</td>
</tr>
<tr>
<td>Android</td>
<td>132</td>
<td>23.2%</td>
</tr>
<tr>
<td>iPhone</td>
<td>97</td>
<td>17.0%</td>
</tr>
<tr>
<td>Tablet</td>
<td>117</td>
<td>20.6%</td>
</tr>
<tr>
<td>Android</td>
<td>48</td>
<td>8.4%</td>
</tr>
<tr>
<td>iPad</td>
<td>69</td>
<td>12.1%</td>
</tr>
<tr>
<td>Computer</td>
<td>223</td>
<td>39.2%</td>
</tr>
<tr>
<td>With mouse</td>
<td>163</td>
<td>28.6%</td>
</tr>
<tr>
<td>With trackpad</td>
<td>51</td>
<td>9.0%</td>
</tr>
<tr>
<td>With touchscreen</td>
<td>7</td>
<td>1.2%</td>
</tr>
<tr>
<td>With trackball</td>
<td>2</td>
<td>0.4%</td>
</tr>
</tbody>
</table>

Amounts are from completer participants (N = 569).

Materials

Face stimuli

There were 143 models in the IMPACT face stimuli, sourced from three stimulus sets. I selected faces that had greater than a 60% hit rate of correct recognition of the emotion being portrayed, which led to 52 disgust and 52 smiling faces from the Umea University Database of Facial Expressions (Samuelsson, Jarnvik, Henningsson, Andersson, & Carlbring, 2012), 73 disgust and 79 smiling faces from the NimStim Set of Facial Expressions (Tottenham et al., 2009), and 51 disgust and 51 smiling faces from the Karolinska Directed Emotional Faces set (KDEF; Goeleven, De Raedt, Leyman, & Verschuere, 2008; Lundqvist, Flykt, & Öhman, 1998). I cropped the NimStim and KDEF images to similar dimensions and face size to match the Umea set, and I performed brightness and contrast enhancement on the KDEF images. The Flanker-Faces task used 16 disgust and 16 neutral faces from the Umea set. The Trained trials used eight disgust faces that also appeared in IMPACT (with three times the frequency for these faces, and their corresponding smiling faces, than for other faces), along with the same models’ eight neutral faces. These eight models’ smiling faces appeared in IMPACT. The Untrained trials involved eight disgust and eight neutral faces from the Umea set not used in IMPACT.
**IMPACT game training**

The game was implemented by our software developer, Howard Braham, using JavaScript, PHP, and MySQL technologies. The design and gameplay of the IMPACT-Positive condition are described above under the heading, “Description of IMPACT active training.” IMPACT-Threat was identical to IMPACT-Positive, except that the tappable stimuli were the disgust faces, and the untappable stimuli were the smiling faces. IMPACT-Undirected differed from the other conditions in that all faces were tappable. Face images had a height of 100 pixels on computers and were a comparable size on tablets and smartphones (using device-independent pixel specification). On tablets and smartphones, the game occupied the full screen, and participants had to hold devices vertically, as the game paused and instructed them to rotate anytime a device was oriented horizontally. On computers, participants could select the window size, but the game maintained an 8:7 pixel ratio of height to width for the playing area, with empty black space on either side horizontally if the window width exceeded this ratio.

Participants completed 15 one-minute rounds of training, with a pause in between rounds requiring a tap to continue. At the start of each round, a reminder of the basic instructions appeared (e.g., “Tap the smiling faces; ignore the disgusted faces” for IMPACT-Positive). Between rounds, any remaining faces were cleared, and the game adjusted its difficulty, as described under the heading, “Motivation by appropriate challenge.”

**Flanker-Arrows attentional measurement task**

The Flanker-Arrows task was based on the Eriksen flanker task (Eriksen & Eriksen, 1974) and the executive control aspect of the attentional network test (Fan, McCandliss, Sommer, Raz, & Posner, 2002). Each trial began with a 500 ms fixation cross. Next, five small arrows appeared in a row, consisting of a central arrow and four flankers. These remained on screen until participants responded to indicate the direction, left or right, of the central target arrow. The flankers’ direction was congruent to the central arrow in half the trials and incongruent in the remaining trials. The difference in mean response time (RT) to congruent versus incongruent trials yielded a score indexing general AC.
Participants using touchscreen devices employed large touch buttons marked “left” and “right” to specify the central arrow’s direction, whereas participants using computers had no on-screen buttons and instead used the arrow keys on the keyboard to respond. The task did not advance until participants responded. Instructions asked participants to respond “as quickly and accurately as possible.” In the 24 practice trials, feedback of “Correct” or “Incorrect” appeared after a response. In the 128 main trials following the practice trials, there was no feedback, hence the inter-trial interval (1000 ms) immediately followed each response.

**Flanker-Faces attentional measurement task**

The Flanker-Faces task was based on past emotional flanker tasks (Moser, Huppert, Duval, & Simons, 2008; Ochsner, Hughes, Robertson, Cooper, & Gabrieli, 2009). Following a 500 ms fixation cross, three faces (from different models) appeared in a row, each portraying a neutral or disgust expression. I chose neutral rather than smiling faces in order to isolate the effects of game training on threat processing while excluding the effects on processing of positive stimuli. Using smiling and disgust faces would match gameplay more directly to measurement; however, doing so would conflate measurement of any training effect of AB towards the tapped stimuli with the training away from untapped stimuli. I wished to use the Flanker-Faces task to investigate effects only on one of these components at a time, to reduce the number of alternative theoretical explanations for how training caused measurement differences.

The two flanker faces’ expression was congruent to the central face in half the trials and incongruent in the remainder. All participants (regardless of device type) saw two large on-screen buttons: a left-side button bearing a neutral face emoticon and a right-side button bearing an emoticon resembling a disgust face. Participants with touchscreen devices responded by tapping the buttons to indicate the central face’s expression, whereas participants on computers used the left and right arrow keys to specify the emotion. As in the Flanker-Arrows task, participants completed 24 practice trials with
feedback about their responses, and in the 128 main trials, the inter-trial interval (1000 ms) immediately followed a response.

In order to separately investigate emotional AC with stimuli used during training (deemed “Trained”) and with previously-unseen stimuli (deemed “Untrained”), I created trials of both types in the Flanker-Faces task. The task included 16 models displaying neutral and disgust expressions, all from the Umea stimulus set. Eight of these models were Untrained and did not appear in the IMPACT game. The other eight were Trained, as the disgust faces of these models were used in IMPACT (though the neutral faces of these models were not used in IMPACT). In order to ensure that participants were familiar with the Trained models’ disgust faces in particular (among the 143 used in IMPACT), these eight models’ faces (both their smiling and disgust faces) appeared with three times the frequency of others during IMPACT.

Speech anticipation stressor

To induce an anxiety reaction in participants, instructions informed participants that they would need to deliver a three-minute audio-recorded speech (Mansell, Clark, Ehlers, & Chen, 1999; Waugh, Panage, Mendes, & Gotlib, 2010). The subsequent screen notified participants of the topic (“Why are you a good friend?”) and gave them 60 s to prepare, showing a countdown from 60 to 0, after which the next screen arrived.

Social Interaction Anxiety Scale (SIAS)

The SIAS (Mattick & Clarke, 1998) is a 20-item measure of social anxiety with strong psychometric properties that may be administered online (Hedman et al., 2010). The three reverse-scored items helped to induce participants to pay close attention to question wording throughout the items (Rodebaugh, Woods, & Heimberg, 2007).
Penn State Worry Questionnaire (PSWQ)

The PSWQ (Meyer et al., 1990) is a 16-item measure of worry that assesses general anxiety unrelated to the social domain. It has strong psychometric properties and may be administered online (Zlomke, 2009). Its five reverse-scored items served to encourage participants to attend carefully to the question wording throughout the experiment, rather than filling out responses without reading. Also, the internal consistency results provide a test of whether participants did read carefully enough to notice the reversals.

Visual Analogue Scale of Anxiety (VAS-Anxiety)

A horizontal bar with a label of “Not at all anxious” on the left and “Extremely anxious” appeared on screen in the Qualtrics survey platform, with instructions to “mark any point along the line to rate how much anxiety you feel at this moment.” A slider notch started at the far left end of the bar. Clicking (on computers) or tapping (on touchscreen devices) anywhere on the bar caused the notch to move to that location, and it could be further adjusted before pressing a “Continue” button to advance to the next screen.

Procedure

First, prospective participants read a description of the study on its MTurk listing. The instructions asked them to complete the survey parts of the study on a computer and switch to a touchscreen device, if available, for the IMPACT game and flanker task parts of the protocol. Only those who had not previously participated in the study were allowed to continue to the Qualtrics survey platform to begin. Via Qualtrics, participants agreed to the consent form. They did not provide their names or email addresses at any point in the protocol. Next, participants responded to demographic questions and self-report questionnaires (SIAS, PSWQ, and PRCS), after which they reported what type of device they would use for the IMPACT game and flanker tasks.
Then, participants completed the VAS-Anxiety and, leaving their Qualtrics window open, continued in a new browser window to a website which randomly assigned them to IMPACT-Positive, IMPACT-Threat, or IMPACT-Undirected training and provided detailed instructions for playing their assigned variant. The instructions explained the type of faces to be tapped or clicked, the point incentives, the rounds, and the adaptive difficulty system. For participants using the same device for both the game and flanker parts as for Qualtrics, a hyperlink brought them straight to the game website. For participants using two different devices (typically a touchscreen device and a computer), the website gave each a unique URL to type into their game devices to begin IMPACT. After 15 one-minute rounds of IMPACT, Flanker-Arrows and Flanker-Faces tasks followed in random order.

After the flanker tasks, participants received a password to enter to continue in the Qualtrics survey back on their original device. They completed a VAS-Anxiety, followed by the speech anticipation stressor, then another VAS-Anxiety. Finally, they received notice that they would not need to deliver any speech, followed by debriefing questions and a full debriefing form. Participants received payment via MTurk within 24 hours.

Data Analyses

I used R 3.1 (R Development Core Team, 2012) for all analyses, except for mixed ANOVA analyses, which I conducted in SPSS 18 (SPSS Inc, 2009).

Self-report measures reliability analyses

I calculated Cronbach’s alphas for the SIAS and PSWQ scales as administered in this sample.

Flanker tasks reliability analyses

I calculated internal consistency estimates for all Flanker task scores using a Monte Carlo process to re-select trials for 1,000 split-half reliability estimates, to increase accuracy and stability (Enock,
Robinaugh, Reese, & McNally, 2012) over traditional methods. I used the Spearman-Brown correction to adjust the estimates for doubled measure length, to compensate for halving.

**Flanker-Arrows analyses**

The sole score calculated for the Flanker-Arrows task was the Conflict score, indexing general AC, where a higher Conflict score indicates lower AC. I calculated this score for each participant by subtracting the mean RT of congruent trials from the mean RT of incongruent trials.

I applied a one-sample t-test to these scores against 0, collapsed across condition, to verify that the expected Conflict effect did occur. To test for group differences, I used a One-Way ANOVA.

**Flanker-Faces analyses**

I calculated several scores for the Flanker-Faces task, in which each trial presented a set of three neutral (N) or threat (T) faces, in arrangements of NNN, TTT, NTN, or TNT. The Conflict score, indexing emotional AC (a higher Conflict score indicates lower emotional AC), was calculated as the mean RT of congruent trials (NNN or TTT) subtracted from the mean RT of incongruent trials (NTN or TNT). The Conflict Ignoring Threat (CIT) score, indexing the emotional AC ability to ignore threat flankers in the task, was a Conflict score calculated within TNT and NNN trials only. The Conflict Ignoring Neutral (CIN) score, indexing the emotional AC ability to ignore neutral flankers, was a Conflict score within NTN and TTT trials only. The AB score was calculated by subtracting the mean RT of trials with a threat central target (TTT or NTN) from the mean RT of trials with a neutral central target (NNN or TNT). The AB score represents the relative response latency of trials with threat targets compared to trials with neutral targets, regardless of the flanker valence.

Since the Flanker-Faces stimuli were divided into Trained stimuli and Untrained stimuli, I calculated subscores for each of the above four scores within Trained and Untrained stimuli, in addition to the overall scores, which collapse across Trained and Untrained stimuli.
For each score, I applied a one-sample t-test against 0, collapsed across condition, to verify that
the expected Conflict effect did occur. Then, to test for group differences, I used Welch’s (equal variances
not assumed) One-Way ANOVA. To follow up on significant group differences identified by the
ANOVA, I used pairwise Welch’s t-tests among the three groups.

VAS-Anxiety analyses

To examine the change in VAS-Anxiety scores over the three time points among the three
training conditions, I conducted a 3 Group (between-subjects factor: IMPACT-Positive, IMPACT-Threat,
IMPACT-Undirected) X 3 Time (within-subjects factor: pre-training, post-training, post-stressor) Mixed
ANOVA. I applied Greenhouse-Geisser corrections due to the data’s failing Mauchly’s Test of Sphericity
(p < .001). To further explore the main effect of time, I conducted Mixed ANOVAs for the first two time
points as 3 Group X 2 Time (within-subjects factor: pre-training, post-training) and the latter two time
points as 3 Group X 2 Time (within-subjects factor: post-training, post-stressor).

Results

Self-report measures reliability

In this sample, Cronbach’s alphas were $\alpha = .96$ for SIAS and $\alpha = .97$ for PSWQ.

Flanker tasks data reduction

Prior to analyses, I performed data reduction on trials from the completer sample, which already
included an accuracy cutoff requiring at least 80% accuracy for each participant’s trials. I used the same
steps as for dot-probe assessment trials in Enock et al. (2014), with cutoffs for outlier RTs of 1,500 ms for
high RTs and 200 ms for low RTs. These cutoffs were determined by Enock et al. (2014), prior to data
analyses, based on a combination of previous data reduction practices used in dot-probe research. For the
present study, I also tested alternative data reduction methods such as using no cutoff, using only a 1,700
ms cutoff (derived from Fan et al.’s (2002) task’s automatic screen advancement if participants did not
respond within 1,700 ms), and using median RTs to obviate the need for a cutoff. The alternative
approaches did not greatly affect the pattern of results, and I determined that adhering to the method previously used for dot-probe allowed for reducing noise due to outliers in the data while maintaining consistency.

Data reduction yielded the following results for Flanker-Arrows task trials: Beginning with 72,833 trials, I first removed inaccurate trials (1.2%), then from the remainder removed trials with RTs less than 200 ms (0.02%) or greater than 1,500 ms (1.6%). From remaining trials, I calculated M and SD for each session and removed trials with RTs more than 2 SDs above (4.1%) or 2 SDs below (0.2%) M.

Similarly, for the Flanker-Faces trials: Beginning with 72,829 trials, I first removed inaccurate trials (3.3%), then from the remainder removed trials with RTs less than 200 ms (0.02%) or greater than 1,500 ms (5.6%). From remaining trials, I calculated M and SD for each session and removed trials with RTs more than 2 SDs above (4.9%) or 2 SDs below (0.1%) M.

Flanker-Arrows task

Table 2 presents descriptive statistics of task RTs by trial type, and Table 3 presents descriptive statistics of Conflict scores by condition, as well as results of the One-Way ANOVA test on these scores, which found no significant differences among conditions (see Table 3).

Flanker tasks reliability

Reliability estimates for all Flanker task scores appear in Table 3.
Table 2. Reaction times in Flanker-Arrows and Flanker-Faces tasks, by trial type.

<table>
<thead>
<tr>
<th></th>
<th>M (ms)</th>
<th>SD (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Flanker-Arrows trial type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>576.0</td>
<td>130.9</td>
</tr>
<tr>
<td>Congruent</td>
<td>547.8</td>
<td>126.9</td>
</tr>
<tr>
<td>Incongruent</td>
<td>605.6</td>
<td>128.4</td>
</tr>
<tr>
<td><strong>Flanker-Faces trial type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>763.3</td>
<td>177.3</td>
</tr>
<tr>
<td>Congruent</td>
<td>760.7</td>
<td>177.1</td>
</tr>
<tr>
<td>Incongruent</td>
<td>766.0</td>
<td>177.4</td>
</tr>
<tr>
<td>Neutral target</td>
<td>772.0</td>
<td>177.0</td>
</tr>
<tr>
<td>Congruent</td>
<td>768.8</td>
<td>177.2</td>
</tr>
<tr>
<td>Incongruent</td>
<td>775.3</td>
<td>176.8</td>
</tr>
<tr>
<td>Disgust target</td>
<td>754.3</td>
<td>177.1</td>
</tr>
<tr>
<td>Congruent</td>
<td>752.2</td>
<td>176.6</td>
</tr>
<tr>
<td>Incongruent</td>
<td>756.4</td>
<td>177.6</td>
</tr>
</tbody>
</table>

RTs are from completer sample accurate trials only, after data reduction.
Table 3. Results from Flanker-Arrows and Flanker-Faces task scores.

<table>
<thead>
<tr>
<th></th>
<th>IMPACT-Positive</th>
<th>IMPACT-Threat</th>
<th>IMPACT-Undirected</th>
<th>Reliability</th>
<th>One-Way ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td><strong>Flanker-Arrows</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conflict</td>
<td>56.0</td>
<td>26.2</td>
<td>61.6</td>
<td>27.3</td>
<td>60.0</td>
</tr>
<tr>
<td><strong>Flanker-Faces</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conflict</td>
<td>4.1</td>
<td>24.8</td>
<td>5.9</td>
<td>26.6</td>
<td>7.8</td>
</tr>
<tr>
<td>Overall</td>
<td>0.9</td>
<td>35.9</td>
<td>4.9</td>
<td>39.4</td>
<td>8.8</td>
</tr>
<tr>
<td>Trained</td>
<td>7.5</td>
<td>36.6</td>
<td>6.8</td>
<td>37.6</td>
<td>6.6</td>
</tr>
<tr>
<td>Untrained</td>
<td>4.1</td>
<td>24.8</td>
<td>5.9</td>
<td>26.6</td>
<td>7.8</td>
</tr>
<tr>
<td>Overall</td>
<td>4.1</td>
<td>24.8</td>
<td>5.9</td>
<td>26.6</td>
<td>7.8</td>
</tr>
<tr>
<td>Trained</td>
<td>0.9</td>
<td>35.9</td>
<td>4.9</td>
<td>39.4</td>
<td>8.8</td>
</tr>
<tr>
<td>Untrained</td>
<td>7.5</td>
<td>36.6</td>
<td>6.8</td>
<td>37.6</td>
<td>6.6</td>
</tr>
<tr>
<td><strong>CIT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>3.7</td>
<td>35.2</td>
<td>9.4</td>
<td>31.8</td>
<td>7.7</td>
</tr>
<tr>
<td>Trained</td>
<td>4.8</td>
<td>48.6</td>
<td>6.4</td>
<td>47.6</td>
<td>8.5</td>
</tr>
<tr>
<td>Untrained</td>
<td>3.2</td>
<td>54.0</td>
<td>11.9</td>
<td>51.2</td>
<td>6.3</td>
</tr>
<tr>
<td><strong>CIN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>4.3</td>
<td>39.0</td>
<td>2.1</td>
<td>42.9</td>
<td>7.2</td>
</tr>
<tr>
<td>Trained</td>
<td>-4.9</td>
<td>55.8</td>
<td>2.7</td>
<td>61.4</td>
<td>8.1</td>
</tr>
<tr>
<td>Untrained</td>
<td>12.2</td>
<td>59.3</td>
<td>1.3</td>
<td>57.2</td>
<td>6.0</td>
</tr>
<tr>
<td><strong>AB</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>15.4</td>
<td>43.4</td>
<td>32.0</td>
<td>52.3</td>
<td>7.9</td>
</tr>
<tr>
<td>Trained</td>
<td>2.6</td>
<td>49.6</td>
<td>22.3</td>
<td>58.9</td>
<td>-5.6</td>
</tr>
<tr>
<td>Untrained</td>
<td>28.0</td>
<td>51.1</td>
<td>41.3</td>
<td>62.1</td>
<td>21.5</td>
</tr>
</tbody>
</table>

* p < .01

The left-hand column shows names of task scores and whether the score is drawn from overall trials (collapsed across Trained and Untrained stimuli), Trained stimuli (where the disgust stimuli were seen during training), or Untrained stimuli (not seen during training). M and SD are in ms. r is a Spearman-Brown corrected split-half reliability. df_{denom} is the denominator degrees of freedom for One-Way ANOVA tests, and the numerator degrees of freedom is 2 for all tests.
Figure 1. Flanker-Arrows Conflict scores by training condition.

Flanker-Faces task

As with Flanker-Arrows results, Table 2 and Table 3 present information about task RTs and task scores. In the sample overall, one-sample t-tests against 0 showed significant conflict effects for Conflict (t(568) = 5.50, p < .001, d = 0.23), CIT (t(568) = 4.79, p < .001, d = 0.20), and CIN (t(568) = 2.72, p = .0067, d = 0.11) scores. One-Way ANOVA tests revealed no differences among conditions for Conflict, CIT, and CIN scores (see Table 3).

AB scores were significantly greater than 0, indicating an AB toward threat versus neutral stimuli, in each of IMPACT-Positive (t(208) = 5.14, p < .001, d = 0.36), IMPACT-Threat (t(171) = 8.02, p < .001, d = 0.61), and IMPACT-Undirected (t(187) = 2.38, p = .018, d = 0.17).
One-Way ANOVA showed significant differences among conditions for all three versions of the AB score (overall, Trained, and Untrained), so I conducted pairwise Welch’s t-tests to compare the conditions. The pattern of results was similar when considering scores overall, Trained, and Untrained stimuli. For AB scores overall, IMPACT-Positive did not significantly differ from IMPACT-Undirected, $t(385.9) = 1.68, p = .093, d = 0.17$; IMPACT-Threat showed higher scores than IMPACT-Undirected, $t(340.5) = 4.64, p < .001, d = 0.49$; and IMPACT-Threat showed higher scores than IMPACT-Positive, $t(332.0) = 3.32, p = .0010, d = 0.35$. For AB scores from Trained stimuli, IMPACT-Positive did not significantly differ from IMPACT-Undirected, $t(388.9) = 1.62, p = .10, d = 0.16$; IMPACT-Threat showed higher scores than IMPACT-Undirected, $t(338.7) = 4.79, p < .001, d = 0.51$; and IMPACT-Threat showed higher scores than IMPACT-Positive, $t(335.4) = 3.48, p < .001, d = 0.36$. For AB scores from Untrained stimuli, IMPACT-Positive did not significantly differ from IMPACT-Undirected, $t(382.5) = 1.20, p = .23, d = 0.12$; IMPACT-Threat showed higher scores than IMPACT-Undirected, $t(343.1) = 3.18, p = .0016, d = 0.34$; and IMPACT-Threat showed higher scores than IMPACT-Positive, $t(330.2) = 2.25, p = .025, d = 0.24$. 
Figure 2. Flanker-Faces Conflict scores by training condition.
Figure 3. Flanker-Faces AB scores by training condition.
**VAS-Anxiety**

In the 3 Group X 3 Time Mixed ANOVA on VAS-Anxiety scores, the main effect of time was significant (F(1.74, 987.49) = 513.29, p < .001), the main effect of group was not significant (F(2, 566) = 0.125, p = .88), and the Group X Time interaction was not significant (F(3.49, 987.49) = 0.16, p = .94). Figure 4 displays the data visually.

Figure 4. VAS-Anxiety scores by training condition over time.
To explore the significant main effect of time, I conducted Mixed ANOVAs by using each pair of adjacent time points. In the 3 Group X 2 Time Mixed ANOVA on scores from the first two time points, the main effect of time was significant ($F(1, 566) = 14.95, p < .001$), the main effect of group was not significant ($F(2, 566) = 0.28, p = .76$), and the Group X Time interaction was not significant ($F(2, 566) = 0.161, p = .85$). The main effect of time indicates a small ($d = 0.14$) but significant ($t(568) = -3.93, p < .001$) decrease in scores in the sample from pre-training ($M = 25.8, SD = 26.5$) to post-training ($M = 22.2, SD = 25.4$).

In the 3 Group X 2 Time Mixed ANOVA on scores from the latter two time points, the main effect of time was significant ($F(2, 566) = 715.98, p < .001$), the main effect of group was not significant ($F(2, 566) = 0.05, p = .95$), and the Group X Time interaction was not significant ($F(2, 566) = 0.15, p = .86$). The main effect of time indicates a large ($d = 1.12$) and significant ($t(568) = 26.90, p < .001$) increase in scores in the sample from post-training ($M = 22.2, SD = 25.4$) to post-stressor ($M = 56.4, SD = 34.8$).

**User experience**

In response to the question, “How enjoyable was playing the Face Game?” participants responded along a 0 (“Not at all”) to 10 (“Extremely”) discrete scale, resulting in $M = 5.85$, $SD = 2.53$, median = 6. A One-Way ANOVA revealed no differences among the training groups, $F(2, 373.5) = 0.50, p = .61$.

**Discussion**

**Overview**

In a large online experiment, I tested the effects of three variants of IMPACT game training on attention and anxiety reactivity. The alternative training conditions did not cause group differences in measures of general AC or emotional AC, but they did lead to differences in AB, showing that task transfer of effects is possible from a fluid, enjoyable game to an RT measurement task. The anxiety-provoking stressor induced a rapid rise in anxiety, but no differences emerged among the training conditions.
Attentional bias effects

The Flanker-Faces AB score showed strong reliability. The AB score’s reliability estimate of \( r = .53 \) was well above those of dot-probe task, as can be seen in the internal consistency estimates given in the first relevant rs reported in each article: Ataya et al. (2012; \( r = .03 \)), Enock et al. (2014; \( r = -.05 \)), McNally et al. (2013; \( r = .07 \)) Schmukle (2005; \( r = -.15 \)), Staugaard (2009; \( r = .10 \)), and Waechter et al. (2013; \( r = .06 \)). Future studies comparing this approach’s reliability directly to dot-probe would be useful.

Training affected AB scores. IMPACT-Threat induced a significantly greater bias toward threat than either of the two other conditions, confirming the proposition of Hypothesis 6. Thus, the game’s design goal of training engagement with a certain valence of faces induced a bias—tested in a very different task—towards these types of faces. This was true for both Trained and Untrained stimuli. IMPACT-Positive did not induce a bias away from threat relative to IMPACT-Undirected, however, a result which does not support Hypothesis 7. Thus, the game’s design goal of training the ability to ignore a certain valence of faces did not produce detectable effects on AB.

Based on AB scores’ being greater than zero within all three training groups, each group showed significant AB toward threat. Speculatively, this effect may have existed prior to training. Or, perhaps interacting with threat stimuli, across the varying ways in which participants did this in the three conditions, could have induced AB, for example by activating threat processing and sensitizing participants to the perception of threat. Moser et al.’s (2008) data showed a different pattern, where participants were faster to respond to central targets with happy or surprise expressions (classified as reassuring) than to those with an angry or disgust expression (which the authors classified as threatening). Comparisons to their results are limited because of the different stimuli used and also because they used the same model for target and flankers in any given trial. One might expect pre-existing AB in the present sample, given the elevated social anxiety, with mean SIAS scores in this sample’s being between half and one standard deviation higher than in some other community and undergraduate samples (Carleton, Collimore, & Asmundson, 2007; Heimberg, Mueller, Holt, Hope, & Liebowitz, 1992). However, AB
scores in this study did not correlate with social anxiety self-report scores, so this explanation is not supported. Pre-training assessment in future studies would help determine whether all IMPACT training variants affected AB, whereas the present data establish only that the IMPACT-Threat training induced a greater AB toward threat than did the other variants.

**Attentional bias measurement approach**

The present approach to assessing AB differs from the visual probe approach. Instead of employing a probe to assess where participants attend, the Flanker-Faces task’s AB score draws on responses to faces that are always centrally presented. In contexts of dot-probe and spatial cueing tasks, researchers have considered AB to potentially represent two visuospatial attentional processes: preferential engagement and delayed disengagement of attention (Clarke et al., 2011; Fox et al., 2001; Koster, Crombez, Verschuere, & De Houwer, 2004; Mogg, Holmes, Garner, & Bradley, 2008). Adding to this landscape, Mogg et al. (2008) investigated the effects of response slowing in the presence of threat stimuli, finding that taking a third aspect into account, response slowing, actually changed the interpretation of score results from evidence of a delayed disengagement from threat to evidence of preferential engagement toward threat. Mogg et al.’s (2008) findings demonstrate the uncertainty inherent in translating observed task scores into inferred attentional shifts that participants make during a task. It has proven difficult to know exactly how task scores result from participants’ actual visuospatial attentional patterns.

In the Flanker-Faces task, the AB measurement does not require an account of engagement or disengagement of visuospatial attention. Here, higher AB scores represent speeded responses when identifying threat targets versus neutral ones. Thus, the results show that IMPACT-Threat trained participants become more biased than the other groups in recognizing and confirming the threat targets versus neutral targets. I have designated this as an AB, as it may result from a greater ease of focusing and maintaining attention to a centrally presented threat stimulus compared to a centrally presented neutral stimulus. Alternatively, it could be conceptualized as an “identification and response bias,” as participants
become faster to identify and respond to threat targets compared to neutral targets. Since it is still a bias of threat versus neutral stimuli, it could be termed an AB, but since it is not a spatial attention bias as dot-probe scores represent, it may warrant a different designation. In either case, the distinct effect of training, transferred across tasks, coupled with the substantial reliability make this bias score worthy of further research. It is important to explore attentional phenomena with a diversity of tasks and not equate the phenomena (AB) with task scores (in such varied tasks such as dot-probe, Flanker, spatial cueing, or emotional Stroop). Having additional tasks and scores may be useful to the development of AB theory.

Attentional control effects

The training groups showed no differences in measures of AC or emotional AC. In Flanker-Arrows, Conflict score reliability was substantial, suggesting that the task did elicit the attentional conflict via flanking arrows as intended. The assessment appears to have measured attention satisfactorily, although the training did not affect AC, thus not supporting the proposition of Hypothesis 1.

In Flanker-Faces, all emotional AC-related scores showed near-zero reliability, suggesting that they did not index substantial individual differences in emotional AC. There was some signal, as scores significantly differed from zero, showing greater RTs to incongruent trials than congruent ones. Yet, overall, it appears that this effect was too weak to produce useful measurement of emotional AC. The propositions of Hypotheses 2 and 3 were therefore not supported. It remains an open possibility that superior measurement of emotional AC could have detected effects.

Anxiety effects

The speech anticipation stressor did produce a marked increase in state anxiety in all groups. It is noteworthy that this type of online stressor was effective, contributing to the evidence base for feasibility of stressor-based online experiments. There were no group differences, however, thus not supporting the propositions of Hypotheses 4 and 5.
Limitations

The present study has limitations. The present sample, not selected for anxiety, could differ from a clinical population with social anxiety disorder, who may be a more appropriate population (though less obtainable at scale) to target for treatment development. The nature of the sample, consisting of MTurk workers, and the online protocol, create a different experimental context from laboratory studies. Online studies have the advantage of scale but have other disadvantages, such as variations in device size and visual angle of screen viewing. Compliance with task instructions would likely have been greater in a laboratory study, and participants would have had fewer distractions. It is conceivable that remote participants could even be multitasking, thus reducing potency of training and increasing measurement error. The reliability estimate for the Flanker-Arrows Conflict score was lower in the present study than in past laboratory studies (e.g., Fan et al., 2002; McNally et al., 2013).

The choice of attentional measurement tasks involves limitations. The measures of emotional AC showed only a weak effect of flanker congruency, and the scores showed near-zero reliability. Thus, it is possible that emotional AC differences could have been present, but that this measure was inadequate to detect them. Future uses of this task could benefit from using the same model for the central target and flankers, to increase the conflict effect, as was done by Moser et al. (2008). I chose to use three different models in each trial for the ecological validity, since people in real-world social situations do not simultaneously see the same person showing multiple expressions, but it may be more important to strengthen measurement reliability. I chose to omit the dot-probe task, due to its history of poor reliability and the need for a brief protocol that MTurk participants would be willing to complete. The AB score assessed in this study showed stronger reliability than typical dot-probe tasks, yet using it makes comparison to existing literature more difficult. Also, it may have been helpful to include positive stimuli in the Flanker-Faces task. In this experiment, I used only neutral and threat stimuli, which was useful to keep the length of the protocol short and minimize dropout, but including all three valences would have allowed for better understanding of attentional changes due to training. Furthermore, in both training and
measurement, disgust faces were the only threat stimuli used. In future research, other threat stimuli should be attempted.

Lastly, the single-session experimental design included only 15 min of training. It is possible that a greater amount of training is needed to increase AC, and the efficacy of multi-session training with IMPACT remains untested.

Conclusions

The present study was the first experimental test of IMPACT. Rather than conduct a small study aimed at testing feasibility, I established feasibility as part of a randomized experiment, delivering the game training across smartphone, tablet, and computer platforms in a web app to a large sample of remote MTurk participants. In doing so, I sought to produce empirical evidence as early as possible that the novel training approach can work. As recommended by Enock & McNally (2013), I bypassed a stage of an open trial without a control group or a comparison of training against no training, instead harnessing low-cost online research methods to accelerate into a more rigorous comparison of three variants of training. Another rigorous choice was employing attentional measurement tasks that were rather different from the training paradigm, in contrast to using the same method for training and measurement of change, as is common in dot-probe ABM research. Achieving task transfer can be difficult: Owen et al. (2010), for example, assigned 11,430 participants to perform various training tasks aimed at improving cognitive function over a six-week period, finding that while participants improved dramatically on the tasks on which they trained, benefits did not generalize even to measurement tasks that used similar cognitive functions.

In the present study, IMPACT training did cause attentional change depending on training variants, documented by the Flanker-Faces AB scores. Unfortunately, the measures of AC and anxiety reactivity yielded no group differences. One insight stemming from these findings is that it is possible to reduce attention to threat without affecting anxiety proneness (as also occurred in Reese, McNally,
Najmi, & Amir, 2010). Further investigation must clarify what changes in attentional responding toward threat may be irrelevant to anxiety versus which are inextricably linked.

Future studies could further probe the changes in attentional responding induced by IMPACT variants. For example, Flanker-Faces AB scores documented a bias toward threat following IMPACT-Threat training versus the other groups. For the IMPACT-Positive group, which tapped positive instead of threat faces, it is quite possible that a positivity bias was induced, but positive stimuli for measurement would be needed to test for this. Future research could test potential additional effects. For example, since a bias in identifying and responding to faces of a certain valence is trained, perhaps other forms of response bias could be identified given the right measures. A lexical decision task involving neutral, threat, and positive words could test whether participants are quicker to identify and respond to words of the valence towards which IMPACT trained.

Ultimately, further research will establish IMPACT’s usefulness, or lack thereof, for experimental manipulation and for anxiety treatment. Regardless of the empirical results for this approach, the pursuit of novel interventions, manipulations, and measures, remains a crucial research endeavor. For both the study of threat processing and the treatment of anxiety disorders, the dot-probe paradigm has shown itself to be useful at times yet unreliable, a catalyst for research but also perhaps a starting point that needlessly narrows the scope of training options. Improvements to dot-probe are certainly useful, and we have seen valuable efforts both to increase its measurement ability for more fine-grained aspects of engagement and disengagement (Clarke et al., 2011) and its usability for training with gamification. Dennis and O’Toole (2014) created a gamified dot-probe task and found that one session of active training reduced anxiety reactivity to a stressor relative to placebo training. The IMPACT project represents a greater departure from existing methods, abandoning the notion of repeated reaction-time trials in favor of a fluid process that continuously engages trainees in desired attentional shifts.
Additional testing of the IMPACT paradigm is warranted. Variations should be explored, given that the first version of an idea is likely not optimal. For example, an attentional flexibility training could be attempted, to train participants to attend at will to either threat or positive faces, by having rounds alternate between IMPACT-Positive and IMPACT-Threat. Or, the game could be simplified by presenting only one type of stimuli, either threat or positive faces, to establish whether interacting with one type of stimuli without the other—testing the effects of merely the presence of and interaction with emotional stimuli, rather than preferential attention—affects AB, AC, or anxiety reactivity. Use of word stimuli in IMPACT instead of faces could also be attempted, enabling research into inducing a bias toward any type of words presented, relative to other words. The possibilities are vast, for this and other potential training approaches. The field of attentional training and computerized treatment research requires continual innovations, and I hope to observe researchers pioneering a multitude of diverse training approaches in the near future.
References


Enock, P. M., & McNally, R. J. (2010). Feasibility and efficacy of attention bias modification via iPhone and other handheld devices to reduce social anxiety and worry. Unpublished manuscript. Department of Psychology, Harvard University, Cambridge, MA.


Appendix A: Speech anticipation stressor

The next part of the procedure is a verbal task: You will give a 3-minute audio-recorded speech on a topic that we will announce on the following page.

You will first have 60 seconds to prepare your speech. Then, we will proceed through the technical setup routine to determine how you will record yourself and ensure that your equipment is prepared to record. Finally, you will deliver the speech, while recording. Your speech will later be judged by our raters for clarity, coherence, and persuasiveness.

[Participant clicks “Continue” button when ready to advance to the next screen, which has the following text and timer.]

In your 3-minute speech, you will speak on the topic of: "Why are you a good friend?" You now have 60 seconds to prepare your speech. Please prepare at this time. The next page will appear in 60 seconds.

[Participant sees a 60-second countdown timer, which begins immediately. When it reaches 0, the next screen (a VAS-Anxiety rating) arrives.]