An affect in its own right: Investigating the case for the independence of social safeness from PA and NA

Benjamin F. Armstrong III

Department of Psychology

McGill University

Montreal, Quebec

September 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctorate of Philosophy

Table of Contents

Abstract

Social safeness is the warm soothing affect system that is theorized to motivate and reinforce social connection and attachment processes (Gilbert, 2014; Gilbert et al., 2008). Together with positive and negative affect (PA and NA), it comprises Gilbert's tripartite model of affect (Gilbert et al., 2008). This three-dimensional model stands in contrast to existing two-dimensional models including the model from which PA and NA were derived (Watson, Clark, & Tellegen, 1988). While social safeness has received growing interest in clinical, affective, and personality research, its independence from and precise relationships with PA and NA are not well established. The aim of the present work was to investigate the precise relationship of social safeness with PA and NA using a variety of designs and populations.

Article One investigated the independence of social safeness from PA and NA in three samples. Sample One was cross-sectional (N = 1889) and recruited from the general community using the online tool Mechanical Turk. Samples Two (N = 93) and Three (N = 99) were both daily diary studies that recruited university students who completed the study both in the lab and online. The independence of social safeness was examined in three ways. First, in each sample, safeness's factorial independence from PA and NA was examined using exploratory and confirmatory factor analysis. Second, social safeness was regressed on both PA and NA in a structural equation model (SEM) in order to determine whether its reliable variance could be fully explained by the two more familiar dimensions of affect. Third, in Samples Two and Three, the unique relationships of social safeness with theoretically relevant constructs were examined at both the between-persons and within-person levels in multi-level SEM. These constructs included perceived stress; perceived, received and given social support; and self-reassurance. Social safeness was found to be factorially distinct in every sample. While PA and

NA did predict substantial variance in social safeness, 35% to 61% depending on the sample, they were by no means able to fully account for safeness's reliable variance. Social safeness had significant bivariate relationships with virtually all stress, support, and reassurance indicators. It showed unique relationships with all indicators at the within-person level and with stress and perceived support at the between-persons level. In summary, social safeness showed strong evidence of independence in all three sets of analyses.

Article Two aimed to further investigate the independence of social safeness from PA and NA while also examining its unique relationships with clinical symptom change. This study was conducted with 92 participants recruited from 18 Mindfulness-Based Stress Reduction (MBSR: Kabat-Zinn, 1982) groups, 65 of whom completed the program. MBSR is a seven-week program of meditation and mindfulness training. The sample had elevated levels of both anxiety and depressive symptoms at intake. This study replicated the analyses from Article One supporting the independence of social safeness from PA and NA both factorially and in terms of variance explained. Social safeness predicted several clinically relevant personality constructs including: mindfulness, self-reassurance, self-criticism, and life satisfaction. It predicted changes in anxiety and depressive symptoms over the course of the group intervention. All analyses were conducted with and without controlling for PA and NA. Social safeness had a bivariate relationship with all study variables except the acceptance subscale of mindfulness, and it uniquely predicted self-reassurance, self-criticism, and life satisfaction. Safeness uniquely predicted change in anxiety but not depressive symptoms.

Taken together, Articles One and Two strongly supported the independence of social safeness from PA and NA. This finding implies that two-factor models are not sufficient to fully describe the varieties of self-reported affect, and a third factor, social safeness, is required. In

addition, the present work provided important empirical support for the relationship of social safeness with support processes related to attachment. The relationships of social safeness with clinical symptomology and clinically relevant personality variables supported both its clinical utility and the conceptual framework underlying Compassion Focused Therapy (Gilbert, 2014).

Résumé

La sécurité interpersonnelle est le système affectif chaleureux et apaisant qui est théorisé pour motiver et renforcer les processus d'attachement et de connexion sociale (Gilbert, 2014; Gilbert et al., 2008). Avec l'affect positif et négatif (AP et AN), il comprend l'ensemble du modèle d'affect tripartite de Gilbert (Gilbert et al., 2008). Ce modèle tridimensionnel contraste avec les modèles bidimensionnels existants, y compris le modèle à partir duquel l'AP et l'AN sont issus. (Watson, Clark et Tellegen, 1988). Alors que la sécurité sociale a reçu un intérêt croissant pour la recherche clinique, affective et de la personnalité, son indépendance et ses relations précises avec l'AP et l'AN ne sont pas bien établies. Le présent travail avait pour objectif d'étudier la relation précise qui existe entre la sécurité interpersonnelle avec l'AP et l'AN en utilisant divers modèles et populations.

L'Article 1 a examiné l'indépendance de la sécurité interpersonnelle de l'AP et l'AN dans trois échantillons. L'Échantillon 1 était transversal (N = 1889) et recruté dans la communauté en utilisant l'outil en ligne Mechanical Turk. Les échantillons 2 (N = 93) et 3 (N = 99) étaient tous deux des études de journal quotidien qui recrutaient des étudiants universitaires qui avaient complété l'étude à la fois en laboratoire et en ligne. L'indépendance de la sécurité interpersonnelle a été examinée de trois manières. Tout d'abord, dans chaque échantillon, l'indépendance factorielle de la sécurité par rapport à l'AP et à l'AN a été examinée à l'aide d'une analyse factorielle exploratoire et confirmatoire. Deuxièmement, la sécurité interpersonnelle a été régressée à la fois sur l'AP et l'AN dans un modèle d'équation structurelle (MES) afin de déterminer si sa variance fiable pouvait être entièrement expliquée par les deux dimensions plus connues de l'affect. Troisièmement, dans les Échantillons 2 et 3, les relations uniques de sécurité interpersonnelle avec des constructions théoriquement pertinentes ont été

examinées aux niveaux inter-personnes et intra-personne dans un MES à plusieurs niveaux. Ces constructions incluaient le stress perçu; le soutien social perçu, reçu et donné; et l'assurance de soi. La sécurité interpersonnelle s'est révélée être, au niveau factoriel, distincte dans chaque échantillon. Bien que PA et NA aient prédit une variance substantielle de la sécurité interpersonnelle, entre 35% et 61% selon l'échantillon, ils n'étaient nullement en mesure de rendre pleinement compte de la variance fiable de la sécurité. La sécurité interpersonnelle entretient des relations à deux variables significatives avec pratiquement tous les indicateurs de stress, de soutien et de réassurance. Elle montrait des relations uniques avec tous les indicateurs intra-personnellement et avec le stress et le soutien perçu au niveau interpersonnel. En résumé, la sécurité interpersonnelle a démontré une forte évidence d'indépendance dans les trois séries d'analyses.

L'Article 2 visait à approfondir l'indépendance de la sécurité interpersonnelle de l'AP et de l'AN, tout en examinant ses relations uniques avec le changement des symptômes cliniques. Cette étude a été menée auprès de 92 participants recrutés dans 18 groupes de Réduction du stress basée sur la pleine conscience (MBSR: Kabat-Zinn, 1982), dont 65 ont terminé le programme. MBSR est un programme de huit semaines de méditation et de formation à la pleine conscience. L'échantillon présentait des niveaux élevés de symptômes anxieux et dépressifs au moment de l'admission. Cette étude reproduit les analyses de l'Article 1 soutenant l'indépendance de la sécurité interpersonnelle de l'AP et de l'AN expliquées de manière factorielle et en termes de variance. La sécurité interpersonnelle prédit plusieurs constructions de la personnalité cliniquement pertinentes, notamment: la pleine conscience, l'auto-assurance, l'autocritique et la satisfaction de vivre. Il a également prédit des changements dans les symptômes d'anxiété et de dépression au cours de l'intervention du groupe. Toutes les analyses

ont été menées avec et sans contrôle pour l’AP et l’AN. La sécurité interpersonnelle avait une relation à deux variables avec toutes les variables de l'étude, à l'exception de la sous-échelle d'acceptation de la pleine conscience, et elle prédisait de manière unique la réassurance, l'autocritique et la satisfaction de la vie. La sécurité prédit de manière unique le changement d'anxiété mais non pas les symptômes dépressifs.

Pris ensemble, les articles un et deux ont fortement soutenu l'indépendance de la sécurité interpersonnelle vis-à-vis de l'AP et de l’AN. Ces résultats impliquent que les modèles à deux facteurs ne sont pas suffisants pour décrire complètement les types d'affect autodéclarés, et qu'un troisième facteur, la sécurité interpersonnelle, est nécessaire. En outre, le présent travail a fourni un support empirique important pour la relation de sécurité interpersonnelle avec les processus de support liés à l'attachement. Les relations de sécurité interpersonnelle avec la symptomologie clinique et les variables de personnalité cliniquement pertinentes ont soutenu son utilité clinique et le cadre conceptuel sous-jacent à la thérapie axée sur la compassion (Gilbert, 2014).

Acknowledgements

I would like to take this opportunity to express my warm thanks to the many people who have supported me in my journey through the PhD program at McGill University. It has been a long road, filled with many ups and downs, and like the many who came before me, I could not have done it without help.

This begins with my supervisor, David Zuroff. David has had a greater impact on my life and way of thinking than anyone else save my immediate family. From the beginning he encouraged, supported, and inspired rather than directing or commanding, and as a result, I have gained not only expertise in the ways and methods of experimental psychology but also a deeper sense of who I am and what I am about in this world. This has come as a result of David's tireless interest, engagement, and support over the past six years. It is with a somewhat heavy heart that I prepare to finish this chapter of my life as a scientist in training. As I move forward, I will always look back on my time with David with fondness and gratitude.

I have had a rich experience of camaraderie and fellowship with my fellow graduate students at McGill University. This begins with the community of lab mates that I was so fortunate to experience, both in the Zuroff Lab directly and in associated labs. Thank you to Daniel Kopala-Sibley, Nicola Hermanto, Clare Foa, Tobey Mandel, Molly Moroz, Kayleigh Clegg, and Gentiana Sadikaj. Many of my ideas and work originated in our extended lab discussions. I would never have gotten here without them. Thanks also to the many assistants and volunteers who came and went over the years: Kimberly Turner, Lia Elbaz, Melody Zhang, Ehime Takahara, Uliana Bilash, Sadya Cousineau, and Elena Heighton. Every one of these wonderful people made a small or not so small contribution to my time here at McGill University, and I am grateful to them. An impossibly large thank you goes to Chantale Bousquet

for helping me with the wide variety of the questions, comments, and bureaucratic problems I encountered (or created for myself) over the years. Thank you to Giovanna LoCascio for helping to keep my academic career on track. And to my clinical cohort who were remarkably supportive and helpful over the many years of our coursework and clinical placements, thank you to Nida Ali, Gilla Shapiro, Esther Yakabov, and Zhen Xu. Thank you to my clinical supervisors and mentors who offered words of wisdom and support as I struggled to navigate a world of conflicting clinical and academic demands: Luisa Cameli, Francine Macinnis, Joe Flanders, Sarah Roberts, Jennifer Russell, Dennis Kalogeropoulos, and Rachel Green.

And thank you to my collaborators, some of whom were with me before I started at McGill and all of whom I hope will be with me long after. Thank you to Drs. Lara Kammrath and Dustin Wood. My life and CV are both richer for our collaborations. And thank you in particular to my collaborator, friend and fellow graduate student Jonas Nitschke. Our hours of meetings, analyses, and chats over drinks made my time in this PhD program immeasurably richer. And to my close friends not yet mentioned, thank you to Sonia Krol, Zach Hershberger and Jeff Anderson. One of the major themes of this dissertation is the importance of warm relationships with others. My relationships with each of them has been particularly warm, and it gladdens my heart to have had them with me.

My final thanks go to my parents, Benjamin and Sharon, and my brother Jonathan. I come from an amazing and supportive family. A primary reason that I have made it as far as I have is because I know that I will always have people who love and care about me to catch me when I fall, to answer my call when I need a friend, and to offer their support and advice whenever necessary. In my 36 years on this planet, I have learned that for those closest to you no effort is too great, and who you are is fundamentally comprised of where you came from and the

relationships that were present there. One of my greatest hopes is that I can one day create a family that is as warm as supportive as the one that I grew up with.

Contribution of Authors

The present dissertation is comprised of two articles and four total studies. I carried out the overall aims of each manuscript under the supervision of Dr. David Zuroff. I also conducted Studies One, Two and Four with the assistance of various lab members. Study Three was conducted by Dr. Zuroff prior to my tenure at McGill University. I was responsible for the creation of all research questions and all analyses in the present dissertation.

Article One is currently being revised for the journal *Personality and Individual Differences* and was coauthored by myself, Jonas Nitschke, Uliana Bilash, and David Zuroff. I designed and secured research ethics board approval for Studies One and Two with input from Jonas Nitschke and David Zuroff. I collected data on those studies with the assistance of Uliana Bilash, Kimberly Turner, and other lab members. Under the supervision of David Zuroff and with support and assistance from Jonas Nitschke, I conducted all statistical analyses and wrote the manuscript.

Article Two is currently being prepared for submission. It was coauthored by myself and David Zuroff. I secured research ethics approval. I designed the study under the supervision of David Zuroff, built the web application through which it was executed, and was responsible for all recruitment and implementation with support from various lab members. I conducted all statistical analyses and interpretation under the guidance and supervision of David Zuroff. I wrote the manuscript with the assistance and feedback of David Zuroff and with editorial feedback from Dr. Sharon Lee Armstrong.

Contribution to Original Knowledge

This dissertation provides a suite of evidence establishing the independence of the affective construct social safeness from existing two-dimensional models of affect, in particular the positive and negative affect (PA-NA) model (Watson et al., 1988). Social safeness was first described by Paul Gilbert (Gilbert et al., 2008). It has important theoretical and empirical relationships with self-compassion, self-criticism, and attachment style (Gilbert, 2014; Gilbert et al., 2009; Kelly, Zuroff, Leybman, & Gilbert, 2012), and it is a primary target for change in Compassion Focused Therapy (Gilbert, 2009, 2014). Even given its theoretical and clinical importance, questions remain about the independence of social safeness from PA and NA. This dissertation provides multiple kinds of methodological evidence for the independence of social safeness (cross-sectional and longitudinal) in student, community, and clinically relevant samples.

Article One explored the independence of social safeness in three samples, one cross-sectional using a community sample recruited from the online-platform Mechanical Turk, and two longitudinal, daily-diary samples drawn from a university student population. The independence of social safeness from PA and NA was assessed in each sample using three methods: factorially using latent variables in SEM, in terms of variance in social safeness explained by PA and NA, also assessed in SEM, and in terms of its unique associations with theoretically relevant outcomes. This work provided broad support in multiple samples and using multiple methods that social safeness is an independent affective dimension distinct from PA and NA. As two of the samples were longitudinal, evidence was also provided at both the between-persons (trait) and within-person (state) levels. Important empirical support was provided for the

relationship of social safeness with social support variables that help to clarify safeness's role in attachment and self-care processes.

Paper Two replicated the independence and outcome analyses of Paper One in the context of a psychological intervention. Paper Two was comprised of one study that followed participants in Mindfulness Based Stress Reduction (MBSR: Kabat-Zinn, 1982) groups over the course of seven weeks. As this sample had elevated levels of both anxiety and depressive symptoms, it contributed to our knowledge of (1) the relationships of social safeness, PA, and NA in a population with clinical features and (2) the ability of social safeness to predict clinical symptom change. To date, social safeness has never been measured as an indicator of clinical change, and this represents an important unique contribution of this paper to the literature on social safeness in particular and clinical psychology outcome research more generally.

General Introduction

"It's like a mother, when the baby is crying, she picks up the baby and she holds the baby tenderly in her arms. Your pain, your anxiety is your baby. You have to take care of it. You have to go back to yourself, to recognize the suffering in you, embrace the suffering, and you get a relief."

- Thich Nhat Hanh

Social safeness is a feeling of warmth and connection to others. It is the sense that all is right with the world and nothing needs to be said or done. Accomplishment can wait until tomorrow. Worry can be set on the shelf for another time. Social safeness is the security of a mother's arms. It is the feeling of being surrounded by friends and family and the sense that one is where one ought to be.

Social safeness is part of Paul Gilbert's tri-partite model of affect (Gilbert et al., 2008). According to this model there are three primary affective systems that guide our behavior. The first is positive affect (PA). PA is our appetitive approach system. It motivates us to seek out resources and opportunities (e.g., hunting for food). It reinforces goal-oriented behavior and provides motivation for accomplishment. PA is associated with the mesolimbic dopaminergic incentive-reward system (Depue, Luciana, Arbisi, Collins, & Leon, 1995; Watson & Naragon, 2009) and provides feelings of enthusiasm, determination, and joy. Contrary to what one might expect, in this model the opposite of high PA is not high NA (negative affect), which is an activated aversion system. Instead the opposite of high PA is the absence of PA: sluggishness or anhedonia that is associated with melancholic depression (Watson, Clark, & Tellegen, 1988; Watson, Gámez, & Simms, 2005).

Negative affect (NA) is our aversive threat system. It motivates us to avoid or confront dangerous or otherwise aversive stimuli (e.g., a stalking tiger). It encompasses states like fear, anger, frustration and anxiety. NA both alerts us to possible threats and also provides motivation and energy for dealing with those threats. It is associated with the cortico-limbic pathway and a variety of neurotransmitter systems including GABA, glutamate, corticotrophin releasing factor (CRF) and others (Mitchell & Phillips, 2007; Panksepp, 1998; Panksepp, Fuchs, & Iacobucci, 2011). In this model the opposite of high NA is not high PA, which as mentioned above, is an activated approach system. Instead, the opposite of high NA is the absence of NA: a state of calm or relaxation.

Gilbert built on the pre-existing two-factor model of PA and NA by adding a third factor, social safeness. Social safeness is associated with being in the presence of close trusted others and produces a sense of connection and security. When one is in a safe place and there is no pressing need for action, social safeness is the system that calms and soothes, promoting rest and recuperation (Gilbert et al., 2008). Social safeness is theorized to be part of an affiliative bonding system that drives the desire to connect with and support others by reinforcing relationships with positive pleasant feelings. It is an important part of the attachment system, driving processes related to parent-child bonding (Gilbert, 2005, 2014; Gilbert et al., 2008). If we draw on the literature related to affiliation/attachment, which is the theoretical evolutionary genesis of the social safeness system, then we have substantial human and animal evidence for specific neurotransmitter systems associated with this system (Depue & Morrone-Strupinsky, 2005; Panksepp, 1998; Panksepp et al., 2011). These systems include oxytocin and vasopressin, which facilitate affiliative approach, and endogenous opioids (mu-opioids and endorphins), which reward and reinforce affiliative contact (Panksepp, 2005).

Gilbert theorized that this third affective system has important implications for self-soothing and therapeutic processes. It is a fundamental part of his Compassion Focused Therapy (CFT: Gilbert, 2009), a third-wave cognitive therapy that emphasizes treating oneself in a warm and compassionate manner. This self-treatment is done in order to self-generate feelings of warmth and safety by self-activating one's social safeness system (Gilbert, 2014). In this way, the CFT therapist assists the client in generating these feelings in the therapy room. Of course, *receiving* compassion and support from others is also theorized to activate the social safeness system. The creation of a warm soothing therapeutic environment may play an important role in therapeutic processes, such as the formation of the therapeutic alliance (Horvath & Luborsky, 1993; Luborsky, 1976) or the creation of the Rogerian conditions of experiencing unconditional regard and understanding (Rogers, 1957).

While PA and NA are both theorized to be affective systems, they also have important personality correlates. That is, the tendency to experience activation in these emotion systems is considered to be trait-like and related to existing personality constructs (Watson et al., 1988). Trait PA is related to extraversion (Watson & Naragon, 2009), and trait NA is related to neuroticism (Watson & Clark, 1985; Watson, Clark, McIntyre, & Hamaker, 1993; Watson & Walker, 1996). Little is known about the personality correlates of social safeness, as this has not been fully investigated. However, links have been shown between social safeness and attachment security (Kelly, Zuroff, Leybman, & Gilbert, 2012).

Social Safeness, Attachment and Other Related Constructs

The social safeness system is conceptually related to but distinguishable from a number of other constructs including: attachment security, perceived support availability, Jaak

Panksepp's CARE and GRIEF systems, social connectedness, and social capital. Here I review these theoretically related constructs.

Attachment security. Social safeness builds on attachment theory (Bowlby, 1980) with an evolutionary account of how the warm soothing feelings of social safeness may have evolved in tandem with the attachment system (Gilbert, 2005). Measurement of attachment security was originally based on Ainsworth's "strange situation" in which a child is separated and then reunited with their mother (Ainsworth & Bell, 1970). Under this paradigm, securely attached children of a certain (young) age will become upset when separated from their mother but will quickly be soothed and return to a calm state after their mother returns. Alternatively, *avoidant* children do not appear distressed by the absence of their mother and avoid contact when she returns, and *anxious-resistant* children become extremely distressed by their mother's absence and have a difficult time being soothed when she returns.

Later work has extended attachment theory to include adult attachment relationships: parents, romantic partners (Hazan & Shaver, 1987), and potentially close friends (Miller & Hoicowitz, 2004). This work has also produced a dimensional model of attachment security whereby attachment security reflects two independent dimensions: *attachment anxiety*, the extent to which one is worried that attachment figures will be available and responsive, and *attachment avoidance*, the extent to which one prefers not to rely on or be close to others (Brennan, Clark, & Shaver, 1998). In the adult attachment literature, attachment security is sometimes viewed as a combination of low attachment anxiety and low attachment avoidance. However, there has been disagreement about this definition as it essentially makes attachment security equivalent to "low insecurity" rather than "an affirmative sense of security in relationships" (Carver, 1997, p. 867).

These concerns led to the creation of the Measure of Attachment Qualities (MAQ: Carver, 1997) which explicitly develops attachment security as a third dimension of attachment style.

An important aspect of attachment style is what Bowlby referred to as a "working model" of the nature of relationships (Bowlby, 1980). These working models are internal representations or schemas that a person holds about oneself and relationships. Under this view, attachment anxiety is characterized by "insecurity about one's own worth and abilities" and "constant worry about being rejected or abandoned" (Gillath, Hart, Noftle, & Stockdale, 2009, p. 162). Attachment avoidance is characterized by "reluctance to trust others," "an emphasis on autonomy and self-reliance," and "a tendency to down-regulate one's own emotions" (Gillath et al., 2009, p. 162). By contrast, attachment security is a "faith in the responsiveness of attachment figures [and] one's own worth and abilities" (Gillath et al., 2009, p. 162). When it comes to state attachment, Gillath et al.'s view is that these anxious, avoidant, and security working models are all present simultaneously and that "levels of attachment anxiety, avoidance, and security are then set as a function of the model or schema that is most strongly activated" (2009, p. 363).

The focus on working models in state attachment provides an important contrast with social safeness. From an affective systems perspective, both the strange situation and adult attachment patterns might represent the tendency or ease with which a person's social safeness system is activated by the presence of an attachment figure, reducing feelings of distress and threat and allowing the person to be soothed (Gilbert et al., 2008). However, as an affective system, social safeness is not defined by the sort of cognitive representations or schemas that characterize attachment style. Where attachment style is a set of beliefs, expectations, rules and scripts for interacting in close relationships (Gillath et al., 2009) that produces feelings of security and connection (or a lack thereof), social safeness is precisely the *feeling* of warmth,

security and connection and the affective machinery that produces it without referencing any cognitive framework that may have been activated.

When one considers the specific items of common attachment measures, the sense of an underlying activated schema is evident. Items from the Experiences in Close Relationships Revised scale (ECR-R: Fraley, Waller, & Brennan, 2000) include belief elements such as: “My partner really understands my needs” and “Sometimes romantic partners change their feelings about me for no apparent reason.” Likewise the MAQ includes items like: “I find others reluctant to get as close as I would like” and “My desire to merge sometimes scares people away” (Carver, 1997), and the State Adult Attachment Measure (SAAM: Gillath et al., 2009) includes such items as: “I feel like I have someone to rely on” or “I feel like I can trust the people who are close to me.” This is not to say that these attachment measures do not have affective elements as well–“I feel secure and close to other people” or “When I’m close to someone, it gives me a sense of comfort about life in general” (SAAM: Gillath et al., 2009)–only that in the conceptualization and measurement of attachment style, the cognitive and affective elements are both present. By comparison, social safeness is measured either by affective adjectives–“safe”, “secure”, “warm”, and “content” (TPAS: Gilbert et al., 2008)–or by general statements that are primarily affective–e.g., “I feel content within my relationships” or “I feel part of something greater than myself” (SSPS: Gilbert et al., 2009).

This brings us to the second key difference between attachment style and social safeness: social safeness is theorized to respond to cues outside of attachment relationships (Gilbert, 2014). While safeness may have evolved to play a role in the formation of attachment bonds, humans can be soothed and feel warmly connected to others who are not attachment figures, even in the broadened sense that includes romantic partners and close friends. Indeed, Gilbert has

stated that the caring and nurturing behaviors associated with social safeness are "not specifically attachment focused. Indeed, we can show care for individuals we may not be attached to or even wish to be attached to. Moreover, care, defined this way can be directed at animals and plants," (Gilbert, 2014, p. 20). Examples of non-attachment safeness and soothing might include being reassured by a boss, colleague or a pet.

Before closing this section on attachment security, it is important to touch on the notion of *felt security*. One way of thinking about attachment is that felt security is the "set goal of the attachment behavioral system" that serves as a "mediator of adaptive behavior" (Sroufe & Waters, 1977, p. 1186). In this view, the attachment system activates and deactivates the feelings and working models associated security in response to safety cues in the environment based on a person's set of internal attachment schemas. Thus proximity-seeking, which is the more classical goal of attachment (J. Bowlby, 1969), is seen as a way to increase felt security. As social safeness is definitionally the in-the-moment feeling of warm security, it would seem reasonable to view social safeness as the affective component of felt security extended beyond the context of attachment to include the warm, secure emotional experiences that occur in a variety of contexts.

Perceived support availability. Barrera (1986, p. 416) describes *perceived support* as "the cognitive appraisal of being reliably connected to others." This definition is consistent with perceived support as the set of beliefs that "one is loved and esteemed by others" and that in periods of stress support will be available when one needs it (Wethington & Kessler, 1986, p. 89). Perceived social support has been shown to buffer the deleterious effects of stress on well-being (Cohen & Wills, 1985) and is linked to lower negative affect and higher well-being (Cohen, 2004). It has a positive relationship with attachment security (Priel & Shamai, 1995) that

is strong enough that it has been suggested that the two constructs stem from the same source (i.e., working models of attachment) or may even be equated (Dunkel-Schetter & Bennett, 1990; Moreira et al., 2003; Sarason et al., 1991). A popular measure of perceived support, the Multidimensional Scale of Perceived Social Support (MPSS: Zimet, Dahlem, Zimet, & Farley, 1988) describes perceived social support as "the subjective assessment of social support adequacy" (Zimet et al., 1988, p.32) and measures it at the trait-level (in general) with such items as "I can count on my friends when things go wrong." and "There is a special person in my life who cares about my feelings." Other instruments have similar items focused on the belief that support is available (Cohen & Hoberman, 1983; Cutrona & Russell, 1987). These measures illustrate the fundamentally cognitive nature of the construct of perceived social support which stands in contrast to the felt affective quality of social safeness. Like attachment security, perceived social support can be measured at the state or momentary level (Crockett & Turan, 2018; Dunkley, Zuroff, & Blankstein, 2003). However, these state measurements parallel existing measures (e.g., Cutrona & Russell, 1987) in focusing on beliefs or appraisals of support availability. In this way, momentary fluctuations in perceived social support function similarly to those in attachment security, with beliefs and ideas being more or less activated moment to moment.

Panksepp's CARE and GRIEF systems. Using neuroscience evidence taken from both human and animal experiments, Jaak Panksepp proposed an affect model composed of seven distinct affective systems (Panksepp, 2005) including: SEEKING, RAGE, FEAR, LUST, GRIEF, PLAY, and CARE. In that model, FEAR and RAGE relate to dealing with aversive stimuli and are conceptually similar to NA. SEEKING, PLAY and perhaps LUST (a link to lust is speculative as no current measures of PA include sexually-oriented content) are approach

systems and conceptually similar to PA. CARE and GRIEF are both oxytocinergic systems related to affiliation and bonding processes in mammals that are conceptually similar to social safeness and attachment. CARE is described as the “maternal nurturance system” and is related to care-giving, particularly in parents toward their offspring (Panksepp, 2010). GRIEF is described as a “separation-distress system” and is related to care-seeking, particularly among young socially dependent mammals (Panksepp, 2010). It should be noted that the conceptualization of GRIEF/separation-distress has gone through a number of iterations. Earlier work labelled it PANIC and/or SADNESS (Panksepp, 1998) although the relationship with attachment processes was still present. The combination of CARE and GRIEF is described as being associated with attachment processes and social bonding (Panksepp, 2005). Like social safeness, CARE and GRIEF could be seen as an affective perspective on attachment phenomena without reference to schemas or working models.

Although conceptually similar, social safeness is different from CARE and GRIEF in several ways. Social safeness is conceptualized primarily as a subjective affective state with some reference to the neural machinery that creates that state. CARE, GRIEF and the other five affective systems in the Panksepp model are primarily conceptualized as behavior-motivation systems and their associated neural machinery. Because this research is done primarily with animals, and ascertaining felt states in animals is problematic, much less attention is given to the felt experience of these states. Instead the primary focus is on observable behaviors and measurable activations in the brain and nervous system. In short, social safeness is a warm-soothing affective state that is theorized to guide behavior. CARE and GRIEF are motivation/behavior systems that appear to have associated affective states in humans and possibly animals.

The affective states associated with CARE and GRIEF are also described somewhat differently from social safeness. GRIEF in particular is described as a distress system (Panksepp, 1998, 2005), which would be more akin to NA than to a lack of safeness. Separation distress is activated in response to separation from an attachment figure, and care-seeking/proximity-seeking is initiated to reduce this distress (Panksepp, 2005). Social safeness is typically described as a positive motivator of proximity and social bonding (Gilbert, 2014; Gilbert et al., 2008), and the underlying system is characterized as moving toward safeness rather than away from distress.

Finally, the CARE and GRIEF systems are narrower in scope than social safeness. They are described in terms of care-giving and care-seeking behaviors, and these behaviors are described only in an attachment context. The felt experience of social safeness is theorized to have arisen in this context, but as mentioned previously, social safeness can be felt in or outside of attachment relationships and is not limited to instances of care-giving or care-seeking. Ultimately, as the underlying neurobiology of the CARE/GRIEF system overlaps with social safeness (Depue & Morrone-Strupinsky, 2005; Gilbert, 2014; Panksepp, 1998), the two perspectives have much in common, though CARE/GRIEF is focused on behavior and described in terms of attachment, whereas social safeness is focused on subjective experience and described in and outside of an attachment context.

Social Connectedness and Need for Belonging/Relatedness. *Social connectedness* is the "short-term experience of belonging and relatedness" (van Bel, Smolders, Ijsselsteijn, & de Kort, 2009, p. 137). *Need for belonging* is the "pervasive drive to form and maintain at least a minimum quantity of lasting, positive, and significant interpersonal relationships" (Baumeister & Leary, 1995, p. 497). In Self-Determination Theory, *need for relatedness* is a fundamental human need for connection and security such that "intrinsic motivation will be more likely to

flourish in contexts characterized by a sense of security and relatedness" (Ryan & Deci, 2000, p. 71). Sidney Blatt (2007, p. 496) characterized achieving relatedness ("the development of increasingly mature, intimate, mutually satisfying, reciprocal interpersonal relationships"), along with achieving positive self-definition, as one of the two fundamental developmental tasks confronting human beings, with both necessary in order to be well adjusted.

These constructs all relate to a need for connection, usually to other humans though animals or religion can sometimes fill this need as well. Need for belonging and relatedness are both characterized as needs, with social connectedness perhaps being the feeling that is engendered by these needs being met. Social safeness differs from the need for belonging/relatedness in that it is conceptualized as an affective system rather than a need. It may be activated in certain contexts (i.e., a need being met), but it is not identical with those contexts. In the psychodynamic model, interpersonal relatedness is a "fundamental psychological dimension," and this dimension is viewed as a part of personality development (Blatt, 2007). This idea is also quite different from the conceptualization of social safeness as an affective system and felt state. Social safeness is distinguished from social connectedness in that the more general sense of belonging may or may not include the warm security that is the hallmark of social safeness. For example, playing soccer in a local league may create a sense of belonging even while the dominant feeling is enthusiasm or even a desire to compete rather than warm security.

Social Capital. *Social capital* is the value or worth of social networks. Originally it was a measure of the interconnected social networks in communities (Jacobs, 1961). Later the expression was used to refer to the resources that a person has available to them through "durable obligations arising from feelings of gratitude, respect, and friendship or from the institutionally

guaranteed rights derived from membership in a family, a class, or a school." (Nahapiet & Ghoshal, 1998, p. 243). That is, social capital refers to the social resources a person has available and can call upon. In this sense it is similar to perceived support but applies more broadly to a wider range of activities including business and political situations, and in fact it is primarily used to describe the capacity for social/political action (Nahapiet & Ghoshal, 1998). As social capital is described as the value of a person's social resources, it is quite different from social safeness and the constructs reviewed above. It has neither cognitive nor affective components and is primarily a measure of the help and support a person is capable of mobilizing toward a goal or end.

Summary. Social safeness shares some similarities, conceptually and in its origins, with several other psychological constructs. Most notably, it is theorized to be an integral part of the attachment system (Gilbert, 2014) and has a strong relationship with *attachment security* at a trait level and with *felt security* at the state or momentary level. Social safeness differs from these constructs both in content and in scope. While attachment constructs involve the activation of "working models" or schemas (Bowlby, 1980), social safeness focuses on mood and the neurobiology that generates it. Where attachment constructs are defined by the context of attachment relationships (broadened to include romantic partners and close friends), social safeness can be felt in any context and triggered by a variety of non-attachment sources. Perceived social support, like attachment security, includes beliefs and schemas at both the trait and momentary levels that are not part of social safeness, which is characterized as an affect. Likewise, social safeness is distinct from the need for belonging/relatedness in that it is not conceptualized as a need but as an affective system, albeit one that may often be activated by the fulfillment of such a need. It is distinct from social connectedness, because while a sense of

belonging *may* activate feelings of social safeness, it is not definitionally the case that it *must*. Social capital is primarily an assessment of the resources a person can mobilize through their social network and is unrelated to any particular affective experience.

One set of constructs did appear to be theoretically and conceptually highly similar to social safeness, Panksepp's CARE/nurturance and GRIEF/separation-distress systems. Like social safeness, CARE and GRIEF are theorized to motivate care-seeking and care-giving processes associated with attachment, and both social safeness and CARE/GRIEF appear related to the same oxytocin-opiate neural circuits (Depue & Morrone-Strupinsky, 2005; Panksepp, 2010). The two perspectives differ in that CARE and GRIEF focus on behavior and are not described outside of attachment contexts whereas social safeness focuses on subjective experience and is not limited to attachment contexts.

PA, NA and the Structure of Affect

There are a variety of different ways of characterizing the structure of affect. Beyond PA and NA, other two-dimensional models include tension and energy (Thayer, 1997), approach and withdrawal (Lang, Bradley, & Cuthbert, 1998), and the valence/arousal model (Russell, 1980). The valence/arousal model is discussed in more detail below. Beyond two dimensions, we have universal emotions based on facial expressions (Ekman & Keltner, 1997), Panksepp's affective neuroscience model (2010), and discrete emotion theories (e.g., Tomkins, 1984). Recent work based on ratings of descriptions of emotions in others found 27 varieties of emotional experience with boundaries between emotion categories described as "fuzzy rather than discrete in nature" (Cowen & Keltner, 2017, p. E7907). Gilbert acknowledges that his tripartite model is a simplification of more complex models with the Panksepp model referenced in particular (Gilbert, 2014). In this dissertation, I focus on positive and negative affect, as they are very

strongly represented in social, personality, and clinical research, and are primary components of Gilbert's tripartite model of affect.

Social safeness stands alongside positive and negative affect as the third part of Gilbert's tripartite model of affect. However, there has been some debate about how positive and negative affect are defined and how they relate to each other. Positive and negative affect have been defined in at least two distinct ways. The first was as part of the circumplex model of affect proposed by Russell (1980). Under this model, the variety of human affective experience is described using two underlying dimensions. The first dimension is *valence*, a dimension ranging from pleasant to unpleasant. The second is *arousal*, a dimension ranging from low to high. Russell proposed these as the underlying dimensions of a circumplex model of affect, suggesting that affective experience exists on a circle encompassing this two-dimensional space (Russell, 1980). See Figure 1. As this model is primarily concerned with the dimensions it defines, i.e., valence and arousal, the terms *positive affect* and *negative affect* do not actually appear in the original conceptualization. However, in later usage, positive affect was often used to refer to pleasant affect, and negative affect was often used to refer to unpleasant affect. In this usage, these terms are synonyms for pleasantness and unpleasantness.

Shortly after the introduction of the valence/arousal model, positive and negative affect were formally defined (Tellegen, 1985; Zevon & Tellegen, 1982) and presented by Watson, Clark, and Tellegen as part of the Positive And Negative Affect Schedule (PANAS: 1988). As part of their effort to characterize the structure of affect, they posited two affective dimensions related to activated emotional states, defining positive affect thusly: "Positive affect is the extent to which a person feels enthusiastic, active, and alert. High PA is a state of high energy, full concentration, and pleasurable engagement, whereas low PA is characterized by sadness and

lethargy" (Watson et al., 1988, p. 1063). They defined negative affect as "a general dimension of subjective distress and unpleasurable engagement that subsumes a variety of aversive mood states, including anger, contempt, disgust, guilt, fear, and nervousness, with low NA being a state of calmness and serenity" (Watson et al., 1988, p. 1063). These definitions were based in part on previous work on the structure of affect that consistently found these two "descriptively bipolar but affectively unipolar dimensions" (Zevon & Tellegen, 1982, p. 112) that were so characterized because there was "emotional arousal (or high affect)" at only one end of the dimension, "whereas the low end of each factor is most clearly and strongly defined by terms reflecting a relative absence of affective involvement (e.g., *calm* and *relaxed* for Negative Affect, *dull* and *sluggish* for Positive Affect)" (Watson & Tellegen, 1985, p. 221). That these dimensions were theorized to be activated states was an important difference from the valence and arousal model of Russell (1980). Watson et al. (1988) further suggested that PA was associated with trait extraversion, and its absence was associated with depression. They suggested that NA was associated with the trait neuroticism and anxiety. See Figure 1.

For many years, these two conceptualizations were thought to be at odds, and a number of studies were conducted to test the supposed bipolarity of positive and negative affect (Feldman Barrett & Russell, 1998; Green, Goldman, & Salovey, 1993; Tellegen, Watson, & Clark, 1999). These studies sought to answer whether positive and negative mood experiences were truly opposite or whether they stemmed from orthogonal (or nearly orthogonal) affective systems. The evidence was mixed, in part because of an ongoing misunderstanding whereby researchers equated the positive and negative *affect* of Watson et al., with the positive and negative *valence* (though often referred to as positive and negative affect) of Russell (1980). The two groups later agreed that their two systems were compatible and that Watson et al.'s PA and

NA are an analytical rotation of Russell's valence and arousal (Russell & Carroll, 1999; Russell, 1999; Watson & Tellegen, 1999). It was suggested that Watson et al.'s PA and NA be referred to as *Positive Activation* and *Negative Activation* to avoid confusion (Tellegen, Watson & Clark, 1999); however, this practice has not been widely adopted by later researchers. While the issue of the bipolarity of PA and NA was largely resolved as a misunderstanding (or mislabeling) of definitions, there were also measurement issues that arose in its testing that are pertinent to the present work.

The Happy-Sad Problem. In testing the proposed bipolarity of positive and negative *valence*, researchers sought to develop pure markers of positive valence and negative valence in order to test the assertion that these markers were (or were not) uncorrelated. The hope was to produce stimuli such that each side of the bipolar dimension could be assessed independently (e.g., "How happy do you feel?" and "How sad do you feel?"). Unfortunately, research employing these markers produced mixed results with happiness and sadness items producing medium-strength correlations that were too strong to support orthogonality and too weak to support bipolarity (Watson & Tellegen, 1999). These results occurred for several reasons but most notably because of the so-called *happy-sad* problem (Watson & Tellegen, 1999). In popular parlance "extremely happy" and "extremely sad" have opposite valences, but "not at all happy" and "not at all sad" do not. Not at all happy implies that one is feeling quite bad, while "not at all sad" implies something closer to neutrality or perhaps mild positivity. These un-matched scalings introduce error and attenuate correlations making measures seem more orthogonal than they really are. See Figure 2.

These considerations are also relevant to the measurement of PA, NA, and social safeness. Because these constructs are unipolar, it is important to understand that the opposite

end of each affective dimension is the absence of that affect rather than an activated state of the opposite valence. Thus, the opposite of high PA is not high NA but rather low PA: sluggishness and lack of motivation. The opposite of high NA is not high PA but rather low NA: calm and serenity (Watson et al., 1988). Likewise, if social safeness is an affective system in the same way as PA and NA, it would also be unipolar with warm feelings of safeness, contentedness and affiliation at the high end, and the absence of those feelings at the low end. If we look at the two measures of social safeness (discussed in detail below) we find that low social safeness implies that one endorses feelings of "not at all […] connected to others" (Gilbert, 2009) or "not at all […] content" (Gilbert, 2008) which is consistent with unipolarity in that these are not activated states of the opposite valence. Some activation at the low end, and thus bipolarity, may exist for the item "safe" as "not at all […] safe" implies some negative activation, although perhaps not to the extent that "not at all safe" can be conflated with "extremely threatened." While Gilbert does not address this directly, he does describe responses to threats as being handled by "other systems" and the absence of social safeness as feeling "separated," "alone," "disconnected," "[that] no-one understands," and/or "unsafe" (Gilbert, 2014). In the present work, social safeness is treated as a unipolar construct as that is what is most consistent with the other two parts of the tripartite model.

Self-Compassion, Self-Reassurance, and Social Mentality Theory

As mentioned earlier, working with the social safeness system is a key aspect of Compassion Focused Therapy (CFT), a system of therapy that emphasizes compassion for others and for the self as a way of regulating emotions. A fundamental precept of this therapeutic model is that social safeness can be self-activated by treating oneself in a compassionate and caring manner. In this way, one treats oneself with *self-compassion* or *self-reassurance*. Self-

compassion is simply the activation of compassion, a behavior and activation pattern that is typically social and other-focused, directed toward the self. Gilbert views self-compassion as an extension of compassion more generally, which he defines as "a deep awareness of the suffering of oneself and of other living beings, coupled with a wish and an effort to relieve it" (Gilbert, 2009b, p. xiii). The parallelism between self-compassion and other-compassion stems from Gilbert's view that compassion is a *social mentality*. Social mentalities are "innate motivation systems which, when activated, organize a range of psychological functions such as attention, emotion, cognition, and behaviour in pursuit of that [social] motive or goal" (Gilbert, 2014, p.11). He goes on to state that although compassion may have evolved to motivate caring behavior directed at others, it can also be directed at the self.

This conceptualization stands in contrast to other definitions of self-compassion, including that proposed by Kristin Neff (2016), whereby self-compassion is the balance between one's ability on the one hand to be kind to oneself, mindful of one's negative emotions, and maintain a larger view of one's position in the greater context of humanity, and on the other hand the extent to which one judges oneself, becomes identified with one's negative emotions, and becomes isolated in the face of negative emotionality. Neff's definition does not appear to parallel other-focused compassion, and in my view may be a recipe for producing self-compassion rather than a definition of the phenomenon per se.

Regardless of the definition, there does not appear to be any debate that cultivating self-compassion is useful because it activates feelings of warmth and caring that are helpful in soothing negative emotions. Gilbert's *self-reassurance*, which is the ability to be self-validating, supportive, and compassionate, refers to one's ability to self-soothe in ways that theoretically target the social safeness system (Gilbert, Clarke, Hempel, Miles, & Irons, 2004). If activating

social safeness is the end-goal of CFT, then self-reassurance is one's ability to treat oneself in ways that facilitate this goal.

Measurement of Social Safeness

Social safeness has been measured using two validated questionnaires, the Social Safeness and Pleasure Scale (SSPS: Gilbert et al., 2009) and the Types of Positive Affect Scale (TPAS: Gilbert et al., 2008). These scales approach the measurement of social safeness differently and therefore have different properties that are worth discussing.

The SSPS is a trait measure of the degree to which a person experiences warm soothing feelings in social situations. It includes statements like, "I feel connected to others," "I feel accepted by people," and "I have a sense of being cared about in the world." These items are given with no timeframe with instructions to describe how one is feeling in "various situations" leading to a sense of describing one's feelings "in general," although this is not explicitly stated. The items were written to describe positive social feelings that are not related to motivation or goal-achievement. Conceptually, these items all relate to a sense of warmth, understanding, and belonging in one's relationships. This seems consistent with the idea that social safeness represents an affective instantiation of attachment security in contexts that range beyond attachment relationships. However, some issues arise with this item content. The first is that while the items do use the word-stem "I feel" or "I have a sense," many are not truly feelings or moods. "I have a sense of being cared about in the world" appears to refer to a cognition or belief. Another item, "I feel part of something greater than myself" is similarly cognitive. Items like this appear to have more in common with measures of belonging, support availability, and attachment security than they do with social safeness as an affective state. While these constructs overlap substantially with social safeness, as outlined above, they are not identical with it. As

such, while this measure likely captures much of the core of trait-level social safeness, it may do so at the expense of missing some important affective qualities of the construct. In addition, this gives the instrument some distinctly trait-like qualities with items like “I feel easily soothed by those around me” being difficult to translate to a daily or momentary context that may not include opportunities for soothing.

The TPAS is an adjective measure of social safeness as well as active and passive positive (valence) affect. For the purposes of this work, only the four social safeness adjective items are relevant: *safe*, *secure*, *warm*, and *content*. These adjectives capture the affective quality of social safeness in a way that the SPSS may miss. They are also very easy and appropriate to administer on a variety of time scales from *right now* to *this week* to *in general*. In this way, the TPAS items represent well the sense in which social safeness is fundamentally an affective system that is active or inactive in the moment. However, the sense that these are fundamentally social states is not as obvious when looking at these items as it is with the SSPS items. In general, the two instruments seem to capture different qualities of social safeness with the SSPS focusing on the larger social context while the TPAS items focus on the more momentary affective qualities. While no study that I am aware of reports the correlation between the SSPS and TPAS, the data collected for Study Two of this dissertation (Armstrong & Zuroff, 2019) indicate a very strong correlation, $r\ (71) = .79$, $p < .001$.

Given the similarity outlined above between social safeness and Panksepp’s CARE and GRIEF systems, it is worth noting that there is a measurement instrument associated with this affective neuroscience approach. The instrument, called the Affective Neuroscience Personality Scales (ANPS: Davis, Panksepp, & Normansell, 2003), gives a series of statements associated with activation of these systems in general. Items related to the CARE system include: “I often

feel the urge to nurture those closest to me," "I am not particularly affectionate," and "I do not especially want people to be emotionally close to me." These items bear some similarity to the SSPS items, with the exception that they focus on giving care and support rather than receiving and being nurtured by it. The GRIEF/separation-distress system is not well-represented in this measure; at the time of the measure's creation, that system was conceptualized more broadly as *SADNESS* and the item content does not include the anxiety and even panic that is associated with separation from loved ones that is part of later iterations of the construct (e.g., "I often feel sad"). Investigating the precise relationships between Panksepp's measure and Gilbert's is beyond the scope of the present work, although formally comparing these measures might bridge the gap between these two disparate parts of affective research.

Empirical Relationships

As reviewed above, social safeness has a theoretical relationship with the attachment system and with attachment security in particular. Gilbert suggests that the development of the attachment system and its attendant affiliative emotions was a turning point in the evolutionary development of mammals beginning over 120 million years ago (Gilbert, 2015). I thus begin this review of the empirical work on social safeness by examining the associations between social safeness and attachment including the assertion that early experiences of warmth from parents in childhood are linked to feelings of social safeness in adulthood (Kelly & Dupasquier, 2016; Marta-Simões & Ferreira, 2018; Matos, Gouveia, & Duarte, 2015; Richter, Gilbert, & McEwan, 2009; Silva, Ferreira, Mendes, & Marta-Simões, 2019) as well as markers of adult attachment, e.g., attachment security and attachment anxiety (Gilbert et al., 2008; Kelly et al., 2012). Building on this is a review of the links between social safeness and how one views and treats the self, including: self-esteem, self-reassurance/self-compassion, self-criticism, and body image

(Marta-Simões & Ferreira, 2018; Mendes, Ferreira, & Trindade, 2018; Pinto, Ferreira, Mendes, & Trindade, 2017). Finally, I provide an overview of the present research on the relationship between social safeness and clinical measures of anxiety, depressive symptoms, and other markers of psychopathology (Alavi et al., 2016; Paul Gilbert et al., 2008; Kelly et al., 2012; Matos et al., 2015).

Links to Attachment. A growing body of research examines the associations between memories of warmth and safeness experienced in childhood and social safeness experienced as an adult. The primary measure used to assess early childhood experience in this literature is the Early Memories of Warmth and Safeness Scale (EMWSS: Richter et al., 2009). The authors of this measure argue that, in line with the suggestions of Gilbert et al. (2003), the *feelings* we recall may be more important for our development than the behaviors or events that triggered them. The EMWSS asks participants to recall the extent to which "I felt cared about," "I knew I could rely on people close to me to console me when I was upset," and "I felt secure and safe." While no studies currently exist linking this instrument directly to attachment style, the item content is highly overlapping with attachment measures that ask about the feeling that "I have someone to rely on" (SAAM: Gillath et al., 2009) or giving "a sense of comfort about life in general" (MAQ: Carver, 1997). Other instruments that assess similar content that appear in this type of research include the EMBU (non-English acronym), which assesses memories of parental rearing behavior (e.g., "abusive", "shaming", "affectionate"–Perris, Jacobsson, Lindström, von Knorring, & Perris, 1980), and the Early Life Events Scale (ELES: Gilbert et al., 2003), which assesses early experiences of threat, submissiveness, and feeling undervalued that have strong negative relationships with the EMWSS (Richter et al., 2009).

The EMWSS has been linked to a variety of adult outcomes. In particular, it has shown a strong relationship with social safeness as an adult (Marta-Simões & Ferreira, 2018; Matos et al., 2015; Richter et al., 2009; Silva et al., 2019) indicating the developmental link between social safeness and early attachment interactions that one would expect if social safeness is part of the attachment system as Gilbert asserts (Gilbert et al., 2008). Reviewing research employing the ELES and EMBU, we find additional associations between adult experiences of social safeness and recalled childhood experiences of rejection, warmth, threat, feeling unvalued, and submissiveness (Richter et al,. 2009; Kelly & Dupasquier, 2016). Recent studies have shown that social safeness mediates the effects of early childhood experiences on current (adult) depressive symptoms (Matos et al., 2015) and the capacity for self-compassion and receiving compassion (Kelly & Dupasquier, 2016), suggesting that the capacity to experience social safeness may be an important mechanism by which childhood experiences of attachment influence adult psychopathology and well-being.

To date, social safeness has only been directly linked to adult attachment style twice. The first time was in the original validation of the TPAS, in which it was shown that safe affect had a medium-sized positive relationship with close (secure) attachment as measured by the Adult Attachment Scale (AAS: Collins & Read, 1990), which uses a three dimensional model of attachment including avoidance, anxiety, and security. Kelly et al. (2012) examined the trait-level relationships between social safeness, as assessed with the SSPS, and attachment style dimensions, as assessed using the Relationship Styles Questionnaire (RSQ: Griffin & Bartholomew, 1994) for close relationships. This instrument can be used to characterize attachment according to the four quadrants created by examining the mental models of self and other that underlie attachment: positive self / positive other (secure), positive self / negative other

(dismissing), negative self / negative other (fearful), negative self / positive other (preoccupied). Social safeness showed a strong positive relationship with secure attachment and a strong negative relationship with fearful attachment. It had small to medium relationships with the other two attachment styles: preoccupied and dismissing. Notably, all these relationships held controlling for PA and NA, and this was to be expected as social safeness has stronger theoretical ties with the attachment system than PA or NA.

Links to self-reassurance, self-criticism, and self-concept. As part of CFT, clients are taught to treat themselves with self-compassion and self-reassurance. (Gilbert, 2009, 2014). Gilbert has suggested that this represents a self-application of the care-giving system to self-soothe and self-activate the security otherwise associated with seeking and receiving care (Gilbert, 2005, 2014). The associations between external and internal caregiving and care-seeking are supported by recent work showing that self-compassion and self-reassurance are predicted by other-focused caregiving and care-seeking behavior (Hermanto & Zuroff, 2016). Correlations between social safeness and self-compassion/self-reassurance were shown cross-sectionally as part of the initial validation of the TPAS (Gilbert et al., 2008). They were also shown in cross-sectional studies looking at disordered eating and body image among undergraduates in Canada (Kelly & Dupasquier, 2016) and in a community sample in Portugal (Mendes et al., 2018). Links between social safeness and self-reassurance have also been found in several other studies that measured both constructs as part of designs serving other aims (Akin & Akin, 2015; Gilbert et al., 2008; Mendes et al., 2018; Petrocchi et al., 2019; Richter et al., 2009).

Social safeness is theorized to have a strong inverse relationship with self-criticism, which Gilbert views as a way of protecting oneself from a perceived threat (Gilbert et al., 2004).

By identifying the pain behind this perceived threat and viewing it compassionately, Gilbert believes that one can "[recode] the emotional memory with new affect processing." (Gilbert, 2014, p.32). In this way, self-reassurance/self-compassion are seen as antidotes or opposites to self-criticism and similar concepts like self-judgment (Gilbert, 2014; Neff, 2016). We would therefore expect to find medium to large relationships between social safeness and self-criticism, and this is exactly what we find. Social safeness showed a medium negative correlation with the self-criticism subscales of the FSCRS in the original validation of the TPAS (Gilbert et al., 2008). These associations were replicated in the Richter et al. study examining the correlations between social safeness and early parental memories (2009), and more recently, in a longitudinal study conducted with college students over the course of a week, Kelly et al. (2012) found a large negative relationship between social safeness and self-criticism at the between-persons level.

Kelly et al. (2012) also showed a large relationship between social safeness and self-esteem, a construct that is highly related to both self-compassion/self-reassurance and self-criticism (Blankstein, Dunkley, & Wilson, 2008; Neff & Vonk, 2009; Zuroff, Moskowitz, Wielgus, Powers, & Franko, 1983). This is part of a larger pattern of associations between social safeness and various measures of self-image. In addition to the aforementioned associations with self-criticism, a construct which includes beliefs about the self (Blatt, D'Afflitti, & Quinlan, 1976; Gilbert et al., 2004), social safeness has been linked to the perception that others view one negatively (Silva et al., 2019). In multiple studies of disordered eating, social safeness has been positively linked to positive body image and negatively linked to disordered eating more generally (Marta-Simões & Ferreira, 2018b; Mendes et al., 2018; Pinto et al., 2017), and in one

very recent study it was related to positive sexual self-identity for members of the lesbian, gay, and bisexual community (Petrocchi et al., 2019).

Links to depression, anxiety, and psychopathology. Given its relevance as a psychotherapeutic target, it is no surprise that social safeness has been measured alongside measures of psychopathology in a number of studies. Given its theoretical connection to attachment and the assertion that it acts as a soothing system more generally, one would expect strong inverse relationships with anxiety disorders, as they are activated states, and social safeness is theorized to de-activate these other affect systems. Moreover, as depression is linked to self-criticism and stress (Blatt & Zuroff, 1992; Hammen, 2005) and as social safeness is theorized to soothe and reduce these phenomena (Gilbert, 2014), one would also expect a negative association between social safeness and depressive disorders. Current research on this topic supports these theoretical assertions. Beginning with initial definition and testing of the construct, Gilbert showed medium-sized negative relationships between social safeness and both depressive and anxiety symptoms as measured by the Depression Anxiety Stress Scales (DASS-21: Lovibond & Lovibond, 1995). These results have been replicated in a number of other studies establishing a clear connection between social safeness and measures of clinical mood dysregulation (Alavi et al., 2016; Gilbert et al., 2009; Kelly et al., 2012; Richter et al., 2009). In addition, Gilbert et al. (2009) measured the relationship between social safeness and markers of disordered mood in a diagnosed bipolar population and found that social safeness was associated with reduced mood instability. Kelly et al. (2012) found negative relationships between trait-level social safeness and depressive symptoms as well as indicators of disordered personality, while controlling for PA and NA. This is the only time that social safeness has been distinguished from PA and NA in predicting clinical symptomology.

Theoretical Predictions for Social Safeness from Two-Dimensional Affect Models

As mentioned previously, Watson et al.'s affect model (1988) asserts that the vast majority of human affective experience can be modeled in approximately two dimensions (Tellegen, 1985), although the existence of additional dimensions was left open. The proponents of the valence and arousal model make an even stronger claim for precisely two dimensions in their assertion that affect exists on a circumplex, whereby human affective states can be plotted on a two dimensional grid (Barrett & Russell, 1998; Posner, Russell & Peterson, 2005; Russell, 1980) comprised of valence on one axis and arousal on the other. Building on this early theoretical work, it has been suggested that the whole notion of affective systems in humans is merely a cognitive elaboration resting on this two-dimensional valence and arousal foundation (Barrett & Russell, 1998; Posner et al., 2005). Of course, this view stands in stark contrast to Panksepp's (2010) affective neuroscience system and to the continuing work of discrete affect researchers who show support for models consisting of, in some cases, dozens of distinct affective states (Cowen & Keltner, 2017).

As the present work is primarily intended to distinguish social safeness from PA and NA, it is important to address just what the alternative predictions would be, both from the PA-NA model and other two-dimensional models such as the valence and arousal circumplex. If two dimensions are sufficient to characterize human affect, we would expect that by measuring just two dimensions we would be able to account for virtually all the reliable variance in social safeness. That is, by measuring PA and NA, we should be able to fully predict a person's standing on social safeness. Likewise, the relationships between social safeness and other constructs should be largely or totally explained by social safeness's relationship with PA and NA. That is, when controlling for PA and NA, social safeness should not be able to meaningfully

predict other constructs. If on the other hand PA and NA are not sufficient to predict a person's standing on social safeness, and/or if social safeness is able to make unique predictions about a person's behavior or standing on other constructs, then this would indicate that social safeness cannot be subsumed by PA or NA, and more than two affective dimensions may be needed. These distinctions would be particularly relevant in those areas where social safeness has established theoretical and empirical links (i.e., attachment and care processes, self-criticism and self-reassurance).

Given that social safeness is part of a tripartite model that posits a system-based view of affect (Gilbert, 2014; Kelly et al., 2012) that could be viewed as incompatible with (or an extension of) existing two-factor models including Watson et al.'s PA and NA and the valence/ and arousal circumplex (Watson, Clark, & Tellegen, 1988), it is somewhat surprising that so little research exists formally analyzing social safeness alongside these two constructs. Only two studies to date have measured social safeness alongside PA and NA. The first, Kelly et al. (2012), is particularly important as it measured these constructs longitudinally yielding both between and within-person relationships. In this study, it was found that social safeness had medium-size associations with social safeness at the between-person (aggregate) level in the expected directions, positive for PA and negative for NA. At the within (daily) level it showed small but significant predicted correlations. As stated above, between-persons relationships with attachment, self-criticism, self-esteem, depression, and markers of disordered personality all held controlling for PA and NA (Kelly et al., 2012). The second study to measure social safeness alongside PA and NA, Kelly and Dupasquier (2016), found only moderate relationships for social safeness with PA and NA cross-sectionally. While moderate relationships are typically

used as evidence for similarity, they are in no way evidence for redundancy and may thus support the distinction of social safeness from PA and NA.

While the aforementioned work distinguishing social safeness from PA and NA is important evidence that social safeness is an independent affective dimension, this evidence could be stronger. There are two ways that the case for the independence of safeness from NA and PA can still be countered. The first argument, and perhaps the more important, is based on effects of shared versus unshared method variance on the magnitude of relations between measures. Kelly et al. (2012) used the Social Safeness and Pleasure Scale (SSPS: Gilbert, 2009), which employs a different question type from the PANAS, general statements rather than affective adjectives. In addition, due to the general nature of some of these statements, it can be difficult to translate to the momentary level, making within-person measurements with it potentially more error-prone. For these reasons, it could be argued that the relationships found by Kelly et al. were not as large as they might have been with an adjective-based measure that more closely parallels the PANAS and can measure momentary states more readily. Stronger relationships of social safeness with PA and NA might cast the independence of social safeness in doubt. The second argument against the evidence for independence of social safeness regards the possibility that the combination of PA and NA might be sufficient to predict a person's standing on social safeness. While previous work has looked at social safeness's relationship with each of PA and NA separately, no study has yet examined whether PA and NA together can fully account for the variance in social safeness, again potentially casting safeness's independence in doubt.

Aims of the Present Work

The overall aim of the present work was to thoroughly test the assertion that social safeness is distinct from PA and NA as defined by Watson et al. (1988). This was done first by examining whether these constructs could be distinguished factorially and second by examining the extent to which they covaried with each other and with other variables in multiple cross-sectional and longitudinal samples.

While the factor structure of the four social safeness items was examined in the original report of the construction and validation of the TPAS (Gilbert, 2008), the safeness items have never been distinguished from the PA and NA items in exploratory factor analysis (EFA) or confirmatory factor analysis (CFA). As such, in each of the four samples in the two articles that follow, we began with EFA and CFA to test the assertion that the safeness items form a separate factor from the PA and NA items. Next, we tested the assertion that a person's standing on social safeness could be fully predicted by PA and NA. This was done using latent variables in SEM to adjust for the effect of reliability on these relationships and was also conducted in all four samples.

After establishing the relationships among safeness, PA, and NA, we examined the unique relationship of social safeness with several theoretically relevant variables controlling for PA and NA. In Article One, Samples One and Two, we examined the relationship of social safeness with perceived stress first cross-sectionally and then longitudinally (measured at the day-level). In Article One, Sample Three, we examined the relationship of social safeness with self-reassurance as well as received, perceived, and given social support. As these constructs have clear theoretical links with social safeness, these represented important tests of whether social safeness could predict unique variance above and beyond PA and NA and thus be considered an independent construct. Finally, in Article Two we examined the relationship of

social safeness with clinically relevant variables (e.g., depressive and anxiety symptoms) in a sample with elevated anxiety and depressive symptomology. This was only the second time that social safeness had been measured in a sample of this type. It was the first time social safeness had been used to predict change in a sample that had undergone a psychological intervention–Mindfulness Based Stress Reduction (MBSR: Kabat-Zinn, 1982).

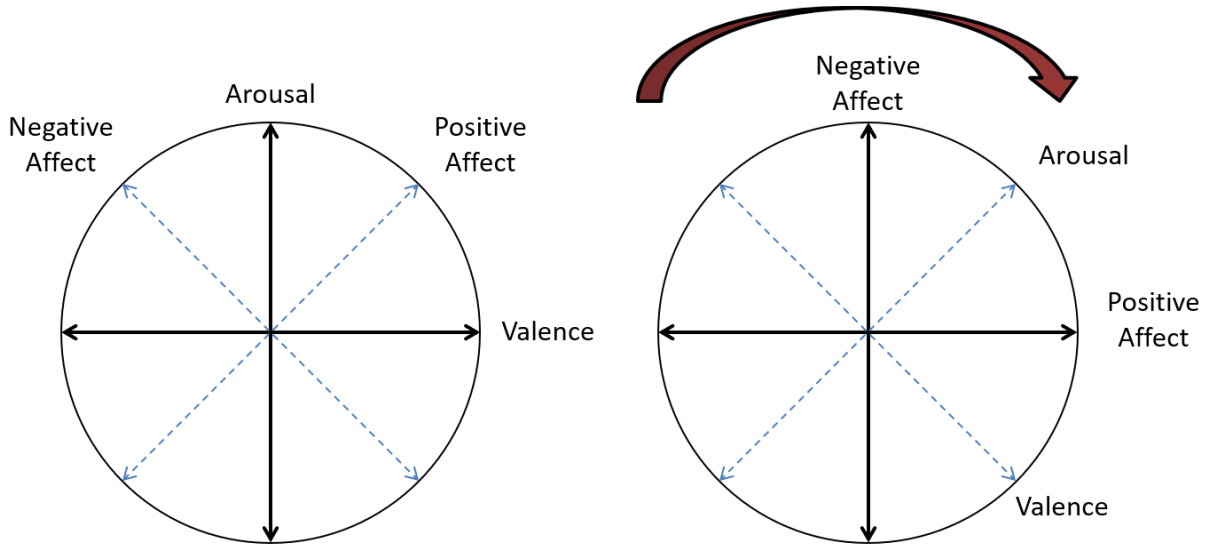

Figure 1. The relationship between valence/arousal and PA/NA affective models. The circle on the left shows the valence and arousal circumplex with valence on the horizontal, arousal on the vertical, positive and negative affect on the diagonals. The circle on the right shows this image rotated 90 degrees yielding the negative and positive affect dimensions on the vertical and horizontal with valence and arousal now on the diagonals.

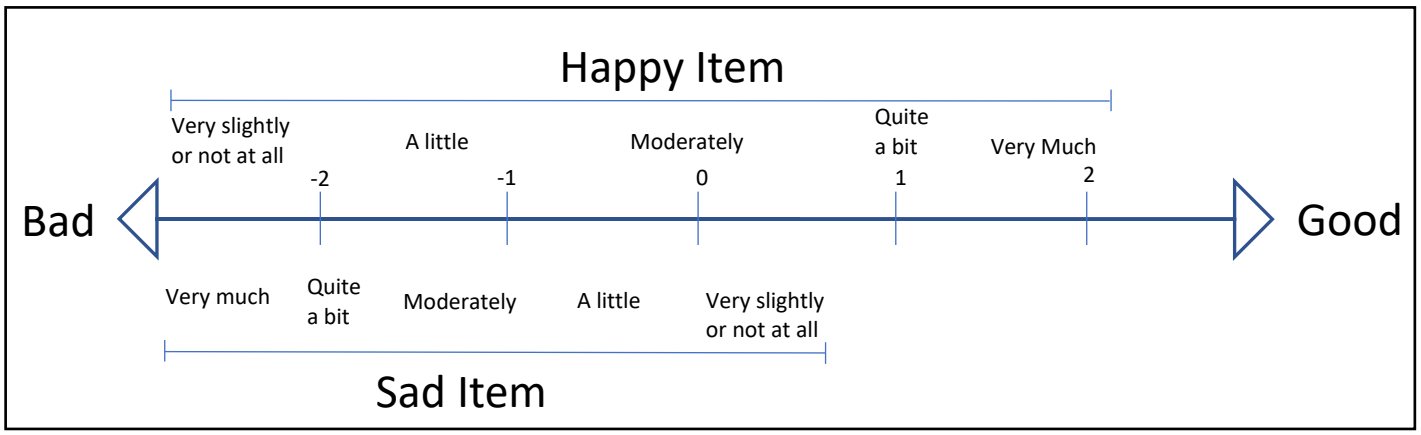

Figure 2. Happy and sad items on bipolar pleasantness-unpleasantness dimension. This is a visual depiction of an IRT analysis conducted by Watson & Tellegen (1999) that showed a non-symmetric overlap between ratings of "How happy are you?" and "How sad are you?" on a latent bipolar pleasantness/unpleasantness dimension here labelled as "Good" and "Bad".

Article One

An Affect in Its Own Right: Investigating the Relationship of Social Safeness With Positive and Negative Affect

Benjamin F. Armstrong III, Jonas P. Nitschke, Uliana Bilash, and David C. Zuroff

McGill University

Armstrong III, B. F., Nitschke, J. P., Bilash, U., & Zuroff, D. C. An affect in its own right: Investigating the relationship of social safeness with positive and negative affect. Submitted to journal, *Personality and Individual Differences*.

Abstract

Many modern theories of affect recognize two fundamental underlying affective dimensions, though the precise dimensions vary. Alongside positive (PA) and negative (NA) affect, social safeness, a warm soothing affect associated with caring and attachment processes, has been proposed as a third fundamental affective dimension (Gilbert et al., 2008). Preliminary work provides some support for the relationships among these three constructs. The present work aims to establish whether social safeness is indeed distinguishable from PA, NA, or their combination. A cross-sectional study consisting of self-report measures of PA, NA, and social safeness was conducted in a sample of 1889 community participants recruited from MTurk. Shared variance among latent variables representing safeness, PA, and NA was examined using correlational and regression analyses. Next, two longitudinal studies of 93 and 99 college students examined the relationships of these three constructs with perceived stress and perceived, received, and given support. Both between-persons and within-person analyses supported the hypothesis that social safeness is an affective dimension in its own right that can be meaningfully distinguished from PA and NA. Social safeness was associated with moderately higher PA and lower NA. It uniquely predicted stress and social support outcomes independently of both PA and NA.

Keywords: social safeness, positive affect, negative affect, social support, stress

An affect in its own right: Investigating the relationship of social safeness with positive and negative affect

Social safeness is a warm, soothing affective state often experienced with close, trusted others. It is part of Gilbert's tripartite model of affect whereby three fundamental affect systems—negative affect (NA), positive affect (PA), and social safeness—evolved as response systems to specific sets of environmental stimuli (Gilbert et al., 2008). According to this view, NA alerts us to aversive and/or threatening stimuli that should be confronted or avoided (e.g., a stalking tiger). PA energizes us to seek out resources or otherwise improve our situation (e.g., hunting for food). Social safeness is theorized to be an affiliative bonding system that gives rise to warm soothing emotions when one is in the presence of close trusted others. Gilbert suggests that in the absence of threat or opportunities for resource seeking, this system creates feelings of peace, safety, and well-being, reducing activation in the other systems and allowing rest and recuperation. Social safeness may relate to attachment processes and contribute to the creation of familial bonds through the shared experience of warm feelings (Gilbert et al., 2008). Hence, the social safeness system is conceptually similar to the *CARE* system proposed by Panksepp (2005) that motivates nurturance and creating familial bonds, particularly maternal ones.

Social safeness has been linked to aspects of psychological well-being (Gilbert et al., 2008; Kelly, Zuroff, Leybman, & Gilbert, 2012), including increased self-compassion (Kelly & Dupasquier, 2016), and reduced indicators of psychological vulnerability and psychopathology, including self-criticism, depressive symptoms, and paranoid traits (Kelly et al., 2012). Although social safeness is socially and clinically important, its precise relationships with other fundamental affective constructs like PA and NA have not been adequately tested, a lacuna the present work addresses.

Social Safeness and the Fundamental Structure of Affect

Gilbert's suggestion that social safeness is an affective construct independent from PA and NA is more than merely an argument for a new affective state. PA and NA are not just individual emotions like *fear* or *enthusiasm*. They are superordinate affective dimensions encompassing individual emotions or moods. PA subsumes emotions such as joviality, self-assurance and attentiveness, and is related to approach motivated behaviours in general. Similarily, NA subsumes emotions such as fear, hostility, guilt, and sadness, and is related to aversion or avoidance motivation in general (Watson & Clark, 1994).

By suggesting that social safeness is distinct from PA and NA, Gilbert is claiming that social safeness is an affective dimension that is potentially at the same hierarchical level as PA and NA (Gilbert et al., 2008). This tripartite model of affect stands in contrast to Tellegen et al.'s (1999) PA and NA model and all other two-dimensional models, including: tension and energy (Thayer, 1997), approach and withdrawal (Lang, Bradley, & Cuthbert, 1998), and valence and arousal (Russell, 1980). These previous models of the structure of affect all suggested two fundamental dimensions. Social safeness is introduced as a third fundamental dimension focused not on approach or avoidance, but on caring and attachment processes.

Defining PA and NA

There are currently two different ways of conceptualizing PA and NA, and importantly, both use the same terms (i.e., PA and NA) but define them differently. One uses them to describe two affective systems (Watson, Clark, & Tellegen, 1988) and the other to describe one dimension of emotional valence (Russell, 1980). This has caused some confusion among affect researchers about whether PA and NA represent two orthogonal constructs or opposite ends of a single bipolar dimension (Russell, 1999; Russell & Carroll, 1999; Watson & Tellegen, 1999). This issue

is particularly relevant to social safeness, as it is defined using systems language, with PA and NA related to approach and avoidance, but was tested using a valence view whereby PA refers to all positively valenced affective states (Gilbert et al., 2008). As our aim is to examine the relationships of PA and NA with social safeness, we will first take a moment to clarify how the terms will be used here.

The first conceptualization of positive and negative affect comes from the valance and arousal literature (Russell, 1980), where *positive affect* refers to positively valanced (i.e., pleasant) emotional states and *negative affect* to negatively valanced (i.e., unpleasant) emotional states. Thus, positive and negative affect are poles of emotional experience with pleasant emotions on one end and unpleasant emotions on the other, each including both high and low energy states. We will refer to this as the *valence view*.

Conversely, Watson et al. (1988) defined positive and negative affect as distinct activated emotion and behavior systems. Thus, positive affect is one's level of activation in a positive direction, and negative affect is one's activation in a negative direction. A key difference from the valence view is that these are activated high-energy states. Low PA and low NA in Watson et al.'s model have the opposite valence from high PA and high NA. For example, high PA in Watson et al.'s model is a state of joy or enthusiasm and is pleasant (i.e., positively valenced). In the valence view, this would also be considered PA. However, low PA in Watson et al.'s model is a state of anhedonia or depression which is unpleasant (i.e., negatively valenced). Because it is unpleasant, in the valence view this would be considered NA. The same thing is true for NA. High NA in Watson et al.'s model is a state of fear or anger and is unpleasant (i.e., negatively valenced). In the valence view, this would also be considered NA. However, low NA in Watson et al.'s model is a state of calm or serenity and is pleasant (i.e., positively valenced). Because it is

pleasant, in the valence view this would be considered PA. So in the low energy ends of the PA and NA dimensions in Watson et al.'s model, the two models disagree. Using the *valence* definitions, PA and NA are opposite poles of a single valence dimension. In Watson et al.'s *dimensional view*, they are orthogonal affective systems. While scholars have debated whether positive and negative affect are *truly* distinct orthogonal dimensions (dimensional view) or opposite ends of a single bipolar valence dimension (valence view), the authors of the two approaches agreed that this issue concerns definitions rather than empirical relationships (Russell & Carroll, 1999; Russell, 1999; Watson & Tellegen, 1999).

In the present work, we will use the dimensional definitions of PA and NA proposed by Watson et al. (1988), these being closest to Gilbert's description of activated behavior systems (Gilbert et al., 2008). It should be noted that the prior work examining the structure of the tripartite model, which includes social safeness, used the Types of Positive Affect Scale (Gilbert et al., 2008). This scale measures two different types of positively *valenced* affect, *activated positive affect* and *relaxed positive affect*, alongside social safeness. We have chosen not to follow this definitional approach and instead to use the dimensional definitions as our aim is to distinguish social safeness from PA and NA according to the theoretical perspective that these are activated systems. Fortunately, valence/arousal and dimensional PA and NA have been shown to be analytical rotations of each other (Watson & Tellegen, 1999), allowing some conversion between them. Thus the prior work on the tripartite model supports the hypothesis that social safeness is distinct from the positive and negative activation dimensions.

We investigated this distinction using a measurement instrument based on the dimensional definition, the Positive And Negative Affect Schedule (PANAS; Watson et al., 1988). Since PA and NA have been demonstrated to be orthogonal on average and to have only a

small correlation in the moment (Watson & Tellegen, 2002), it will also be important to distinguish social safeness from PA and NA for one's general disposition (i.e., between-persons) *and* for the more immediate experience of emotion states (i.e., within-person).

Relations among Social Safeness, PA and NA

While important work linking social safeness to well-being and clinical constructs exists, only two studies have measured the relationship of social safeness with PA and NA using an instrument from the dimensional perspective. Both showed only a modest correlation of social safeness with PA and NA (Kelly & Dupasquier, 2016; Kelly et al., 2012), suggesting that these constructs are indeed distinct. However, the measure they employed, the Social Safeness and Pleasure Scale (Gilbert et al., 2009), uses broad general statements (e.g., "I find it easy to be calmed by people close to me") not well-suited to measuring more immediate momentary emotional states. These statements also differ greatly in format from the adjectives found in the PANAS, suggesting that safeness correlations with NA and PA were likely attenuated. A proper state-level measure of social safeness comparable to the PANAS will make it possible to get a clearer picture of the true relationships among these variables at both the between-persons and within-person levels. As similar instrument formats tend to produce higher correlations regardless of the real relationship between constructs, this test will more rigorously assess whether these constructs are distinct.

The opposing view to Gilbert's would be that social safeness can be largely or totally explained using the two existing affect dimensions. The literature on collinearity and multicollinearity provides guidance for conditions under which variables become redundant with one another. Values of 80% or 90% shared variance have been suggested as thresholds for multicollinearity (Kleinbaum, Kupper, Nizam, & Rosenberg, 2013). Thus, if PA and NA together

accounted for 80-90% of the variance in social safeness it would be clear evidence that safeness is subsumed by PA and/or NA. If they together accounted for much less variance in social safeness, perhaps even less than even a 50% majority, this would be strong evidence of safeness's independence.

We can also examine social safeness's ability to uniquely predict other constructs. Meaningful unique predictions of important outcomes above and beyond PA and NA would further mark social safeness as an independent affective dimension and potentially an affective system in its own right.

The Present Article

In this article, we sought to investigate the relationship of social safeness with PA and NA using a multi-study approach that employed both cross-sectional and longitudinal paradigms and also adjective-based measurement instruments that were as methodologically similar as possible across the three constructs. To test the independence of social safeness as rigorously as possible, we used latent variable models that correct for measurement error and disattenuate correlations.

In the cross-sectional study, we aimed to replicate previous work on zero-order and unique relationships among the three constructs controlling for each other. We hypothesized that social safeness would show medium to large latent associations with both PA and NA separately ($r > .30$ or ~10% of the variance), but not associations so large as to be considered collinear with either construct (i.e., $r < .90$ or ~80% of the variance). In addition, a substantial amount of variance in social safeness was expected to remain unexplained by PA or NA when both are included as predictors. We hypothesized PA and NA together would account for less than 80% of the total variance in social safeness and perhaps less than a simple 50% majority.

Next, we conducted a nine-day daily diary study to examine the relationships among PA, NA, and social safeness at the between-persons (trait) and within-person (state) levels. To be an independent affect system, safeness must be independent from other affective constructs at both the trait (between-persons) and state (within-person) levels. Thus, we predicted that PA, NA and social safeness would be correlated but distinct with a substantial amount of variance unaccounted for by PA and NA both between- and within-person. We also investigated whether social safeness predicted unique variance in an important outcome, perceived stress, controlling for the other two affective dimensions. We hypothesized that social safeness would significantly predict variance in perceived stress above and beyond PA or NA both between and within persons.

Crucially, to the test the external validity of social safeness we conducted a second daily diary study over seven days to examine the relationships of social safeness, PA, and NA with important social support constructs: provided support, received support, and perceived support availability. Safeness is theorized to be the affect system driving attachment and support processes related to providing and seeking care (Gilbert et al., 2008). Feeling safe may lead one to be more willing to provide support or take the risk of seeking it. Thus, we hypothesized that it would show a strong unique relationship with provided *and* received support.

Perceived support availability is associated with a sense of safety and security that has been shown to act as a buffer against stress (Cohen & Wills, 1985; Mikulincer & Shaver, 2009). This is similar enough to the definition of social safeness that one might argue that both terms describe the same construct. We suggest that while related, perceived support availability is a more cognitive construct, while social safeness is a more affective one. This distinction would be consistent with the findings of Kelly et al., (2012) showing a unique relationship between social

safeness and several indicators of psychosocial vulnerability and maladjustment. Similar to those findings, we hypothesized that social safeness and perceived support availability would show a strong relationship but not so strong as to be redundant. We further hypothesized that social safeness would show a strong unique relationship with perceived support availability above and beyond PA and NA.

Method

Participants were recruited as part of three studies employing measures of PA, NA, and social safeness. The first study (*Sample One*) was cross-sectional and conducted entirely online using the online participant recruitment tool Amazon Mechanical Turk (MTurk). Participants needed access to an internet-connected computer or mobile device. *Sample One* consisted of 2223 English-speaking North American adults between 18 and 79 (M = 34.4, SD = 12.0). They were primarily female (N = 1200, 54%), and the majority had completed at least some post-secondary education (87%).

The second study (*Sample Two*) was conducted over nine days, online and in the lab, using a college student sample recruited at [NAME OF UNIVERSITY]. *Sample Two* consisted of 96 participants recruited through online classified ads, Facebook, and flyers. Participants were primarily female (N = 63, 66%) and ranged in age from 18 to 50 (M = 22.60, SD = 4.53).

The third study (*Sample Three*) was conducted over seven days, online and in the lab, with a college student sample of 99 participants recruited at [NAME OF UNIVERSITY] via Facebook, Craigslist and university classified ads. Participants were primarily female (N = 51, 52%) and ranged in age from 18 to 25 (M = 20.5, SD = 1.76).

Participants failing to complete 10% or more of the items in any study measure or taking on average less than one second per item were removed from Samples One and Two. This criterion was selected based on previous work showing that it takes at least one second to

respond thoughtfully to questionnaire items (Wood, Harms, Lowman, & DeSimone, 2017). In total, this reduction represented approximately 15% of Sample One (reduced N = 1889) and 3% percent of Sample Two (reduced N = 93). Approximately 18% of Sample Three was removed for failure to submit responses during the designated time-frame (see below and Zuroff et al., 2016). Given the large size of Sample One, statistical power was deemed sufficient for finding even very small effects. According to recent work on sample size requirements for SEM (Wolf, Harrington, Clark, & Miller, 2013), Samples Two and Three may be slightly underpowered for determining between-person relationships, although this may be offset somewhat by our prediction of moderate to large effect sizes. These samples had 556 (Sample Two) and 690 (Sample Three) observations (approximately 5.98 and 6.97 per participant respectively). This should be sufficient for determining within-person effects. The study protocol for Sample One (REB File #: 250-1214), Sample Two (REB# 152-0816), and Sample Three (REB# 121-0910) were approved by the research ethics board at [NAME OF UNIVERSITY].

Procedure

Participants in Sample One responded to an online advertisement on MTurk. Upon answering the ad, they were directed to an online consent form and given the study instructions. The study took approximately 5-7 minutes, and subjects were paid approximately 20 cents (US) for their time.

Participants in Sample Two were recruited after answering one of several ads posted on Facebook, the [NAME OF UNIVERSITY] campus, or a [NAME OF UNIVERSITY] student website. Upon replying to the ad, they were invited into the lab where they received the consent form and a study description. The intake procedure consisted of study instructions as well as an

initial battery of personality and mood measures taking approximately 40 minutes to complete. Next, each evening, participants filled out online questionnaires about their mood throughout the day. The study also included experience-sampling data collected using participants' smart phones at 4-6 daily time points; however, the experience-sampling data used experimental affect measures not germane to the present work and will not be described further. Participants were compensated up to $28 for their time, depending on their level of participation.

Participants in Sample Three were recruited after answering ads posted on Facebook, Craigslist, or on [NAME OF UNIVERSITY] campus. Upon responding, they were invited into the lab where they were given the consent form, informed about the study, and asked to fill out questionnaires on a lab computer. For the next seven days, they received daily email links to complete online measures of PA, NA, social safeness and several social support related constructs. Measures were completed between 6pm and 4am, or they were considered missed. Participants were compensated $16 for the initial laboratory session, $2 for each completed diary, and a $20 bonus for completing all seven diaries. Sample Three has featured in previously published work (Hermanto et al., 2017; Zuroff et al., 2016).

Measures

Participants in Sample One completed all instruments using the time-frame "right now" and those in Sample Two using "today." Participants in sample three completed all instruments using the time-frame "today." Reliabilities for all measures were calculated using McDonald's omega (McDonald, 1999), which makes fewer assumptions than Cronbach alpha. It has been demonstrated to be a more accurate reliability measure under many circumstances and is particularly well-suited to structural equation models, including hierarchical models (Dunn, Baguley, & Brunsden, 2014; McDonald, 1999; Trizano-Hermosilla & Alvarado, 2016).

Positive and negative affect. Positive and Negative Affect were assessed using the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988). The PANAS is a 20-item measure consisting of two lists of 10 mood adjectives, one for positive affect (e.g., *enthusiastic*) and one for negative affect (e.g., *nervous*). Participants responded using a 5-point Likert scale ranging from 1 (*very slightly or not at all*) to 5 (*extremely*). McDonald's omega was .92 for positive affect items in Sample One. In Sample Two, omega was .80 for positive affect items at the within level and .90 at the between level. In Sample Three, omega was .94 for positive affect items at the within level and .99 at the between level. Omega was .92 for negative affect items in Sample One. In Sample Two, omega was .86 for negative affect items at the within level and .98 at the between level. In Sample Three omega was .88 for negative affect items at the within level and .99 at the between level.

Social safeness. Social safeness was assessed using the Types of Positive Affect Scale (TPAS; Gilbert et al., 2008). TPAS is an 18-item measure consisting of adjectives related to three types of positive affect: active (e.g., *energetic*), relaxed (e.g., *peaceful*), and safe (e.g., *secure*). Participants rated how they felt using a 5-point Likert scale ranging from 0 (*Not characteristic of me*) to 4 (*Very characteristic of me*). Only the four items related to social safeness were relevant to the present work. McDonald's omega for social safeness items was .88 in Sample One. In Sample Two, omega was .77 at the within level and .93 at the between level. In Sample Three, omega was .77 at the within level and .93 at the between level.

Perceived stress. Perceived stress was assessed using the Perceived Stress Scale 4-item version (PSS4; Cohen, Kamarck, Mermelstein, et al., 1994). The PSS4 consists of items related to feelings of stress, overwhelm and loss of control (e.g., *unable to control the important things in your life*). Participants rated how often they felt stress on a 5-point Likert scale ranging from 0

(*never)* to 4 (*very often*). The PSS4 was not administered in Sample One. McDonald's omega in Sample Two was .63 at the within level and .89 at the between level.

Self-reassurance. Self-reassurance was assessed using a shortened 5-item version of the self-reassurance subscale from the Forms of Self Criticising and Self-Reassurance Scale (FSCRS: Gilbert et al., 2004) derived in previous work (Hermanto et al., 2016) by selecting the highest loading items in the original sample. The self-reassurance subscale measures the extent to which participants were able to reassure themselves when facing difficulties or failures (e.g., *I was able to feel lovable and acceptable* and *I encouraged myself for the future*). Participants rated how they felt on a 5-point Likert scale ranging from 0 (*never*) to 4 (*very often*). The FSCRS was only administered at the day-level and only in Sample Three. McDonald's Omega was .99 at the within level and .99 at the between level.

Received, given, and perceived social support. Received, given, and perceived social support were measured using shortened three-item versions of the received and given support subscales of the Social Provisions Scale (SPS; Cutrona, 1989; Cutrona and Russell, 1987). Each subscale asks the participant the extent to which they received, gave or perceived availability (respectively) of each of three social provisions–guidance, tangible assistance, or emotional closeness. Responses were rated on a 7-point Likert scale ranging from 1 (*not at all*) to 7 (*very much*). The three-item versions of the SPS subscales used have been used in previous studies (e.g., Dunkley, Zuroff, & Blankstein, 2003; Kelly, Zuroff, Leybman, & Gilbert, 2012; Zuroff, Sadikaj, Kelly, & Leybman, 2016). Received, given, and perceived support were administered at the day-level and only in Sample Three. For received social support, omega was .55 at the within level and .98 at the between level. For given social support, omega was .60 at the within level

and .98 at the between level. For perceived social support, omega was .74 at the within level and .99 at the between level.

Data Analytic Plan

First, both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were conducted in Sample One using *R* (R Team, 2017), *psych* (Revelle, 2017), and *lavaan* (Yves, 2012) packages to investigate the factor structure of a combined pool of PA, NA, and social safeness items. This allowed investigation of any shared item-level variance in these scales. To conform to best practices, which require EFA and CFA to be conducted in separate samples, Sample One was randomly split into two separate samples maintaining gender parity (Thompson, 2004). Both incremental and absolute fit indices are reported including the comparative fit index (Bentler, 1990) and the root mean-square error of approximation (Steiger & Lind, 1980). As recent recommendations for model-fit cutoff criteria are much stricter than previous guidance (Hu & Bentler, 1999), we will report fit as *excellent* if it meets stricter criteria (CFI $\geq$ .95, RMSEA $\leq$ .06), *good* if it meets previous more liberal criteria (CFI $\geq$ .90, RMSEA $\leq$ .08) and *marginal* if it almost meets the more liberal criteria. Although we report chi-square, we do not use it as a model selection guide as it has been shown to be too conservative for very large samples (Hu & Bentler, 1999). In Samples Two and Three, model fit was assessed with the same criteria and designations using multilevel SEM with *Mplus 8* (Muthén & Muthén, 1998).

Following factor analyses, both observed correlations of composite (sum of items) scores and standardized betas derived using structural equation models (SEM) were calculated in Samples One and Two to examine the relationships of PA, NA, and social safeness. In Sample One, analyses were conducted in the full sample rather than in one of the two random

subsamples. In Samples Two and Three, these relationships were examined at both the between- and within-person levels.

Finally, multilevel SEM analyses were conducted to examine the unique relationships of these variables with perceived stress in Sample Two and for self-reassurance and perceived, received, and given support in Sample Three. All analyses examined relationships at both the between- and within-person levels. Many analyses conducted in Sample Three produced correlation matrices that were not positive definite due to residual variances that were either negative or undefined. When necessary, these residual variances were fixed to zero to resolve this issue. This is noted in the text.

The measurement model for PA and NA in our SEMs used an item parceling strategy previously employed with the PANAS (Dunkley et al., 2017; Dunkley, Ma, Lee, Preacher, & Zuroff, 2014). Recent work has indicated that parceling can have a number of psychometric benefits, including enhanced indicator reliability and the creation of indicators from Likert items suitably continuous for maximum likelihood estimation (Little, Rhemtulla, Gibson, & Schoemann, 2013). Our parceling procedure involved selecting every third item from each subscale leading to two three-item parcels and one four-item parcel for each PANAS subscale. A primary concern in parceling is that the procedure will change or obscure aspects of the factor structure; however, the PANAS has a well-established factor structure at both the between- and within-person levels, meaning that parceling is unlikely to change results for subsequent analyses (Dunkley et al., 2017; Sterba & MacCallum, 2010). As social safeness had only four items, which would have resulted in only one parcel, no parceling procedure was used.

Results: Sample One

Exploratory Factor Analysis

An item-level Principal Axis EFA was conducted in the first half of Sample One using maximum likelihood (ML) estimation and promax rotation, yielding a 3-factor structure. The three factors had eigenvalues of 5.67, 5.54, and 2.67 and explained 24%, 23%, and 11% of the variance respectively. While a fourth factor had an eigenvalue above one (1.56), a scree test (Cattell, 1966) indicated retaining three factors. Parallel analysis (Horn, 1965) suggested retaining a fourth factor; however, the potential factor had only two clear items ("attentive" and "alert") and one item with a substantial cross-loading (.38). Extracting factors with fewer than three items and/or substantial cross-loadings with other factors is commonly used as grounds to exclude additional factors (Ward, 2006). We concluded that the data best supported a three-factor solution. Rotated factor loadings can be seen in Table 1. The three extracted factors mapped very well to the three constructs. All NA items loaded well on Factor One; all PA items loaded well on Factor Two, and all social safeness items loaded well on Factor Three. The PA and NA factors had a small negative correlation, $r(276) = -.18$, $p < .001$. Social safeness had a moderate correlation with both PA, $r(276) = .51$, and NA, $r(276) = -.54$, $p < .001$.

		Factor Loadings		
Subscale	Item	Positive	Negative	Safeness
PA	PANAS2	**0.66**	0.03	-0.14
PA	PANAS4	**0.75**	0.01	-0.11
PA	PANAS8	**0.70**	0.01	0.04
PA	PANAS6	**0.82**	-0.07	0.19
PA	PANAS7	**0.89**	0.04	0.00
PA	PANAS11	**0.55**	-0.15	-0.08
PA	PANAS13	**0.82**	-0.05	0.15
PA	PANAS15	**0.76**	0.05	-0.06
PA	PANAS18	**0.60**	0.10	-0.08

PA	PANAS20	**0.85**	0.08	-0.03
NA	PANAS1	-0.06	**0.69**	0.01
NA	PANAS3	0.13	**0.78**	0.00
NA	PANAS5	0.07	**0.73**	0.09
NA	PANAS9	0.01	**0.86**	-0.01
NA	PANAS10	0.07	**0.74**	0.05
NA	PANAS12	-0.03	**0.54**	0.02
NA	PANAS14	0.09	**0.86**	-0.02
NA	PANAS16	-0.03	**0.74**	-0.05
NA	PANAS17	-0.16	**0.59**	0.01
NA	PANAS19	-0.02	**0.80**	-0.07
Safeness	Safe1	-0.06	-0.01	**0.79**
Safeness	Safe10	0.00	-0.09	**0.89**
Safeness	Safe11	0.08	0.12	**0.78**
Safeness	Safe12	-0.10	0.11	**0.68**

Table 1

Rotated Factor Loadings (Promax) for PA, NA, and Social Safeness Items

Note. Factor loadings (>.40 in bold) are taken from an Exploratory Factor Analysis of all items from the positive and negative affect schedule (PANAS: Watson et al., 1988) and the Type of Positive Affect Scale (TPAS: Gilbert et al., 2008). PA = positive affect, NA = negative affect. Safeness = social safeness.

A CFA was conducted using MLR estimation to confirm the 3-factor model found in EFA using the second half of Sample One and the parceling procedure described previously. This model produced good fit, x^2 (32) = 180.258, CFI = .972, RMSEA = .070. Again, PA and NA factors had a small negative correlation, r (32) = -.16, $p < .001$, and social safeness had a moderate correlation with both PA, r (32) = .50, and NA, r (32) = -.60, $p < .001$.

Associations among PA, NA, and Social Safeness

Bivariate associations. Both observed correlations of composite scores and standardized covariances of latent variables were calculated to determine the bivariate relationships among PA, NA, and social safeness (see Table 2). Consistent with previous work, PA and NA showed a small negative association, and social safeness showed a moderate positive association with PA and a moderate negative association with NA. Latent variable associations were highly similar to observed correlations. Overall, PA accounted for 25% of the variance in social safeness, and NA for 34%. This was consistent with our hypothesis that PA and NA would each account for more than 10% of the variance in social safeness but less than 80%.

	PA	NA	Social Safeness
PA	-	-0.15***	0.50***
NA	-0.15***	-	-0.58***
Social Safeness	0.48***	-0.52***	-

Table 2

Correlations among PA, NA, and Social Safeness in Sample One

Note. Unshaded correlations in the lower triangle are correlations between observed (composite) scores. Shaded correlations in the upper triangle are standardized covariances (i.e., correlations) between the latent versions of the three constructs derived from structural equation modeling. ***$p < .001$.

Partial regression coefficients. The unique effects of PA and NA in predicting social safeness were also calculated using SEM with variances of the latent variables set to 1.0. It was found that the effects of PA and NA in predicting social safeness were largely unique, with PA having a significant positive association, $\beta = .43$, $SE = .020$, $z = -20.7$, $p < .001$, and NA having a significant negative association, $\beta = -.51$, $SE = .034$, z = -20.6, $p < .001$. PA and NA were modestly but significantly associated, $\beta = -.15$, $SE = .014$, z = -6.39, $p < .001$. Together, PA and NA accounted for less than a majority of the variance in social safeness (44%) and did not

approach the 80% cutoff that would indicate that social safeness is redundant with them. These unique associations and the substantial percentage of variance in social safeness not accounted for by PA and NA were taken as evidence that social safeness is distinguishable from both PA and NA.

Results: Samples Two and Three

Between-Persons and Within-Person Associations of PA, NA, and Safeness

Bivariate Associations. As in Sample One, correlations (standardized covariances) of the latent variables were calculated in SEM. As Samples Two and Three were longitudinal, this was done using multilevel SEM and maximum likelihood (MLR) estimation in the statistical program *Mplus* (Muthén & Muthén, 1998). Model fit was good in Sample Two (CFI = .94, TLI = .92, RMSEA = .057) and Sample Three (CFI = .95, TLI = .93, RMSEA = .056). Intraclass correlations for parcels and indicators as well as factor loadings for Sample Two and Sample Three are provided in Tables 3 and 4 respectively. For standardized covariances of the latent variables, see Table 5 and 6.

		Loading		
Variable	Indicator	Between	Within	ICC
PA (parcels)	P1	.89	.86	.53
PA	P2	.95	.86	.54
PA	P3	.87	.63	.63
NA (parcels)	P1	.95	.78	.41
NA	P2	.96	.81	.38
NA	P3	.89	.70	.51
Safeness (items)	I1	.93	.77	.54
Safeness	I2	.96	.63	.52

Safeness	I3	.84	.67	.52
Safeness	I4	.79	.68	.46
PSS (items)	I1	.61	.46	.29
PSS	I2 (reversed)	.95	.50	.38
PSS	I3 (reversed)	.90	.66	.31
PSS	I4	.69	.63	.48

Table 3

Standardized Factor Loadings and Intraclass Correlations for PA, NA, and Safeness in Sample Two

Note: Intraclass correlations and standardized between and within factor loadings from a two-level measurement model conducted in Mplus (Muthén & Muthén, 1998-2017) for all study variables.

		Loading		
Variable	Indicator	Between	Within	ICC
PA (parcels)	P1	.85	.88	.47
PA	P2	.95	.81	.48
PA	P3	.93	.78	.47
NA (parcels)	P1	.90	.79	.53
NA	P2	1.00	.85	.42
NA	P3	.90	.67	.50
Safeness (items)	I1	.93	.64	.40
Safeness	I2	.83	.75	.31
Safeness	I3	.99	.79	.37
Safeness	I4	.81	.72	.44
PS (items)	I1	.94	.63	.53
PS	I2	.93	.50	.62
PS	I3	.98	.55	.57
GS (items)	I1	.88	.62	.29
GS	I2	.83	.64	.39

GS	I3	1.00	.56	.34
RS (items)	I1	.99	.61	.36
RS	I2	.81	.59	.45
RS	I3	1.00	.57	.37
REAS (items)	I1	.81	.73	.48
REAS	I2	.85	.59	.51
REAS	I3	1.00	.56	.33
REAS	I4	.86	.48	.45
REAS	I5	.82	.37	.34

Table 4

Standardized Factor Loadings and Intraclass Correlations for PA, NA, and Safeness in Sample Three

Note: Intraclass correlations and standardized between and within factor loadings from a two-level measurement model conducted in Mplus (Muthén & Muthén, 1998-2017) for all study variables. Loadings are shaded for indicators whose residual variance was fixed to zero to facilitate proper model specification. This was done for all models using these indicators.

	PA	NA	Safeness
PA	x	-0.42***	0.56***
NA	0.04	x	-0.71***
Safeness	0.64***	-0.44***	X

Table 5

Correlations among PA, NA, and Social Safeness in Sample Two

Note. df = 64. Correlations are the standardized covariances calculated using latent variables in a multilevel structural equation model. Unshaded associations in the lower triangle are at the between-persons level. Shaded associations in the upper triangle are at the within-person level. *$p < .05$. **$p < .01$. ***$p < .001$.

	PA	NA	Safeness
PA	x	-0.28***	0.52***
NA	0.10	x	-0.41***
Safeness	0.64***	-0.22+	X

Table 6

Correlations among PA, NA, and Social Safeness in Sample Three

Note. df = 64. Correlations are the standardized covariances calculated using latent variables in a multilevel structural equation model. Unshaded associations in the lower triangle are at the between-persons level. Shaded associations in the upper triangle are at the within-person level. $+p < .10$. $*p < .05$. $**p < .01$. $***p < .001$.

Between-person associations were highly similar in these samples to those in Sample One, excepting that the relationship between PA and NA, which had shown a small negative correlation in Sample One, was here close to zero. As in Sample One and as predicted, the between-persons associations of PA and NA with safeness were far less than the threshold of redundancy of 80% shared variance.

At the within-person level, the relationship of PA and NA was larger than at the between-persons level, showing small to moderate negative correlations. In Sample Two, social safeness showed a small significant positive association with PA and a large significant negative within-person association with NA. In Sample Three, it again showed a small significant association with PA and, in contrast to Sample Two, its association with NA was small and marginal rather than large. Again, these relationships were far below the threshold of redundancy.

Partial regression coefficients. As in Sample One, unique effects of PA and NA in predicting social safeness were calculated by simultaneously regressing social safeness on PA

and NA, this time both between and within persons. Both PA and NA had significant associations with safeness at the between-persons level in Sample Two, β_{pa} = .65, *SE* = .10, $p < .001$; β_{na} = -.45, *se* = .08, $p < .001$, and Sample Three, β_{pa} = .67, *SE* = .10, $p < .001$; β_{na} = -.28, *se* = .09, $p < .001$. These relationships were also significant at the within-person level in Sample Two, β_{pa} = .31, *SE* = .07, $p < .001$; β_{na} = -.57, *SE* = .05, $p < .001$, and Sample Three, β_{pa} = .44, *SE* = .04, $p < .001$; β_{na} = -.29, *SE* = .05, $p < .001$. Together, PA and NA accounted for 61% of the between-persons variance in social safeness in Sample Two, r^2 (64) = .62, *SE* = .14, $p < .001$, and 49% in Sample Three, r^2 (64) = .49, *SE* = .12, $p < .001$. They accounted for 56% of the within-person variance in Sample Two, r^2 (64) = .56, *SE* = .065, $p < .001$, and 35% in Sample Three, r^2 (64) = .35, *SE* = .056, $p < .001$. In both samples, it was clear that while PA and NA had substantial relationships with social safeness, the latter was not redundant with nor reducible to NA and/or PA.

Bivariate and partial associations of social safeness with perceived stress in Sample Two. We calculated the individual and unique contributions of PA, NA and social safeness in predicting perceived social stress in Sample Two. Fit for this model was good (CFI = .93, TLI = .91, RMSEA = .051). Results are available in Table 7 (individual) and Table 8 (unique). All three variables showed large to very large relationships with perceived stress; in particular, the relationship between social safeness and perceived stress was large with nearly 60% shared variance at the within level and 68% at the between level.

	Perceived Stress	
	Between	Within
PA	-.66*	-.66*
NA	.54*	.78*
Safeness	-.83*	-.82*

Table 7

Correlations of PA, NA, and Social Safeness with Perceived Stress in Sample Two

Note. df = 156. Correlations (standardized covariances) were calculated using latent perceived stress, PA, NA, and social safeness in a measurement model using multilevel SEM.
* $p < .001$.

	Perceived Stress	
	Between	Within
PA	-.42*	-.29*
NA	.38*	.38*
Safeness	-.39*	-.39*

Table 8

Standardized Regression Coefficients Predicting Perceived Stress in Sample Two with NA, PA and Social Safeness

Note. df = 156. Regression coefficients are from a multilevel SEM in which latent perceived stress was regressed on latent PA, NA, and social safeness.
*$p < .001$.

When NA, PA, and social safeness were entered in the model as simultaneous predictors of perceived stress at both levels (Table 8), the standardized partial regression coefficients (betas) were moderate, indicating that safeness was a significant predictor of perceived stress, at both levels, even when controlling both NA and PA. These moderate unique associations of social safeness with perceived stress at both the between- and within-person levels, controlling for PA and NA, was taken as additional evidence that social safeness is indeed an affective construct in its own right, distinct from PA and NA.

Bivariate and partial associations of social safeness with given, received, and perceived support as well as self-reassurance in Sample Three. We calculated the individual and unique contributions of PA, NA and social safeness in predicting given, received, and perceived support as well as self-reassurance in Sample Three. Separate models were conducted for each outcome to prevent model misspecification. Results are available in Tables 9 (bivariate) and 10 (unique). Substantial bivariate relationships existed between PA, NA and safeness and the support and reassurance variables (Table 9). At the within level, PA, NA and Safeness all had a large association with self-reassurance. PA had medium associations with perceived, received and given support. NA had small associations with perceived and given support but no association with received support. Safeness had small to medium associations with all three support variables. At the between level, only PA and safeness had associations with self-reassurances, and these associations were large. PA and NA both had small to medium associations with all three support variables, though the association between NA and received support was marginal. Safeness had a large association with perceived support, a small one with received support, and no association with given support.

	PA		NA		Safeness	
	Bet	With	Bet	With	Bet	With
Perceived Support	.25*	.33***	-.39***	-.23***	.72***	.34***
Received Support	.43***	.36***	.20+	.03	.26*	.27***
Given Support	.47***	.45***	.29*	-.15*	.14	.33***
Reassured Self	.72***	.67***	-.16	-.52***	.63***	.68***

Table 9

Correlations of PA, NA, and Social Safeness With Perceived, Received, and Given Support as Well as Self-Reassurance in Sample Three

Note. df = 169. Correlations (standardized covariances) were calculated using latent perceived stress, PA, NA, and social safeness in a measurement model using multilevel SEM in Mplus. Significant correlations are shaded grey.
* $p < .05$. ** $p < .01$. *** $p < .001$. + $.05 < p < .10$

	PA		NA		Safeness	
	Bet	With	Bet	With	Bet	With
Perceived Support	-.37	.20**	-.18	-.09	.93***	.21*
Received Support	.38*	.32***	.17	.19*	.06	.19*
Given Support	.56**	.37***	.19	.02	-.17	.15*
Reassured Self	.61***	.40***	-.19+	-.26***	.19	.36***

Table 10

Standardized Regression Coefficients of PA, NA, and Social Safeness With Perceived, Received, and Given Support as Well as Self-Reassurance in Sample Three

Note. df = 156. Regression coefficients taken from a multilevel SEM in which each variable was regressed on latent PA, NA, and social safeness.
* $p < .05$. ** $p < .01$. *** $p < .001$. + $.05 < p < .10$

When NA, PA, and social safeness were entered in the model as simultaneous predictors of each variable at both levels (Table 10), the standardized partial regression coefficients (betas) associations appeared smaller, indicating that PA, NA, and safeness share some variance in predicting these variables. Nonetheless, safeness had unique associations with each variable at the within-person level and with perceived support at the between-persons level. Importantly, the unique association of safeness with perceived support was very large (r = .93) between-persons, and in this model neither PA nor NA remained significant, indicating safeness and perceived support may be highly overlapping constructs between-persons.

The unique associations of social safeness controlling for PA and NA with all support and self-reassurance variables at the within-person level and perceived support at the between-persons level was taken as additional evidence that social safeness is indeed an affective construct in its own right, distinct from PA and NA.

Discussion

Social safeness is an affective dimension representing soothing affiliative emotions introduced by Paul Gilbert (Gilbert et al., 2008). It is conceived as part of a tripartite model of affect whereby it represents one of three fundamental affective systems: PA, NA, and social safeness. The purpose of the present work was to clarify the relationships among these variables and establish whether social safeness is an affective system in its own right, distinct from the other two. This would represent an important addition to existing two-factor affect models. In order to address this important question, we conducted three studies: one cross-sectional and two longitudinal.

We first established that PA, NA and safeness are factorially distinct in EFA, where items separated cleanly into three factors without substantial cross-loadings. We then examined the relationships among these constructs cross-sectionally and longitudinally. In the cross-sectional sample, PA and NA showed a small association as found in previous work. As predicted, social safeness was related to, but not redundant with PA and NA. Critically, PA and NA together accounted for less than 50% of the variance in social safeness, much less than would be required to conclude that it is redundant with the two more familiar affect dimensions.

This pattern of distinctness among the three affective constructs also appeared in the two longitudinal samples, making this as the first time that all three were compared using state-level instruments of comparable format (affect adjectives). Findings for PA and NA conformed with previous longitudinal work (Watson & Tellegen, 2002) showing no relationship at the between-persons level and a small-to-moderate negative relationship at the within-person level. One's levels of PA and NA are unrelated on average, but in the moment one can affect the other to some degree. Positive and negative affect appear to originate from distinct affective systems that are largely independent (Watson et al., 1988).

As in the cross-sectional sample and as predicted, social safeness had substantial but not redundant relationships with both PA and NA in the two longitudinal studies. The total variance explained in social safeness by PA and NA, while higher than hypothesized in Sample Two, was still well below the threshold for redundancy, both between- and within-person and in both samples. This again suggests social safeness is distinct from PA and NA. One can experience various combinations of emotions like anger/fear (NA), enthusiasm/joy (PA), and security/warmth (social safeness) at least somewhat independently. This independence supports the idea that social safeness is an independent affective system.

In addition to showing that social safeness is not redundant with PA and NA, we also demonstrated its substantial unique associations with perceived stress and several constructs related to social support. Social support predicted unique variance in perceived stress, and as perceived stress could be viewed as very similar to NA, that social safeness predicted unique variance above and beyond NA is important evidence of safeness's independence.

Safeness was also an important unique predictor of perceived, received, and given support at the within level, and perceived support at the between level. As safeness has been theorized to be related to attachment and caring processes, one would expect it to show an important association with support and caring related processes. This was exactly what was found, with social safeness showing at least a small unique association with each support variable at the within level and a very large unique association with perceived stress at the between level. It also showed a medium unique association with self-reassurance. This is relevant to Gilbert's social mentality theory, which asserts that the ability to self-sooth relates to the ability to soothe and care for others, allowing one to self-activate feelings of safeness that would otherwise come from support processes. Here we have presented evidence that self and other caring processes are related to the same affective system, social safeness.

Contrary to theoretical work asserting that social safeness is a system that inhibits both PA and NA leading to periods of rest and recuperation, we found small to moderate bivariate and unique associations between social safeness and PA. This was in addition to the negative relationship between social safeness and NA. If social safeness is instrumental in attachment processes, it may act as a protective factor against stress and an enhancement factor for positive motivation. This would be related to the idea of a "secure base" in attachment theory (Bowlby, 2005; Holmes, 2001), which reduces anxiety and encourages exploration. As the present study

measured these constructs at the day-level rather than in the moment, this idea is speculative and requires further testing. It is also possible one could experience both high levels of safeness and goal-orientation, for example, over the course of a day but not in the same moment. It will be important for future researchers to continue investigating the precise relationship of social safeness with other affective and care-related constructs in the future.

Important limitations of the present work include characteristics of the two samples and the within-person time-frame. The cross-sectional study was conducted online using the recruitment service MTurk. While this service has been shown to produce valid data in previous research (Crump, McDonnell, & Gureckis, 2013; Hauser & Schwarz, 2016), it may not generalize well to other less educated and technologically adept populations. Both the longitudinal samples were comprised of college students in their early 20s and may also have restricted generalizability. Participants in these samples assessed their emotional experience at the day-level rather than at the level of momentary experience. This limits our ability to draw conclusions about the in-the-moment relationships among the affective variables, making our interpretation at the emotional systems level tentative. Other limitations include the exclusive use of self-report measures and the lack of a psychological manipulation. Implicit or physiological measures would allow stronger assertions regarding the systems that may underlie these constructs. Direct manipulation of these constructs would allow causal relationships to be investigated among these dimensions and with other psychological outcomes.

Conclusion and Future Directions

We have presented evidence that social safeness is an important affective dimension at least partially independent from PA and NA. Where previous work has generally found two fundamental dimensions, we have presented evidence for third. This suggests not only that social

safeness is an important affective construct that bears further investigation but also that more work is required on the structure of affect and affective systems more generally. Social safeness is not orthogonal to other fundamental affective dimensions. It does not neatly extend current two-dimensional models into an orthogonal third dimension. This suggests that the structure of affect may be more complex than current theories propose. Social safeness also showed important hypothesized relationships with self and other caring constructs, reinforcing the notion that it is an important part of attachment and caring processes. It will be for future researchers to more fully investigate the precise nature of social safeness, its relationships with other affective constructs, and its place in the overall structure of affect. This may involve measuring social safeness at the level of immediate experience or activating the state directly using an experimental manipulation.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–46.

Bowlby, J. (2005). *A secure base: Clinical applications of attachment theory* (Vol. 393). London and New York: Taylor & Francis.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245–276.

Cohen, S., Kamarck, T., Mermelstein, R., & others. (1994). Perceived stress scale. *Measuring Stress: A Guide for Health and Social Scientists, 10*.

Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological bulletin, 98*(2), 310.

Cutrona, C. E., & Russell, D. W. (1987). The provisions of social relationships and adaptation to stress. *Advances in personal relationships*, *1*(1), 37-67.

Cutrona, C. E. (1989). Ratings of social support by adolescents and adult informants: Degree of correspondence and prediction of depressive symptoms. *Journal of personality and social psychology*, *57*(4), 723.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS One, 8*(3).

Dunkley, D. M., Lewkowski, M., Lee, I. A., Preacher, K. J., Zuroff, D. C., Berg, J.-L., … Westreich, R. (2017). Daily stress, coping, and negative and positive affect in depression: Complex trigger and maintenance patterns. *Behavior Therapy, 48*(3), 349–365.

Dunkley, D. M., Ma, D., Lee, I. A., Preacher, K. J., & Zuroff, D. C. (2014). Advancing complex explanatory conceptualizations of daily negative and positive affect: trigger and maintenance coping action patterns. *Journal of Counseling Psychology, 61*(1), 93-109.

Dunkley, D. M., Zuroff, D. C., & Blankstein, K. R. (2003). Self-critical perfectionism and daily affect: Dispositional and situational influences on stress and coping. *Journal of Personality and Social Psychology, 84*(1), 234–252.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412.

Gilbert, P., Clark, M., Hempel, S., Miles, J.N.V. & Irons, C. (2004). Criticising and reassuring oneself: An exploration of forms, styles and reasons in female students. *British Journal of Clinical Psychology*, *43*, 31-50.

Gilbert, P., McEwan, K., Mitra, R., Franks, L., Richter, A., & Rockliff, H. (2008). Feeling safe and content: A specific affect regulation system? Relationship to depression, anxiety, stress, and self-criticism. *The Journal of Positive Psychology*, *3*(3), 182–191.

Gilbert, P., McEwan, K., Mitra, R., Richter, A., Franks, L., Mills, A., … Gale, C. (2009). An exploration of different types of positive affect in students and patients with bipolar disorder. *Clinical Neuropsychiatry, 6*(4), 135–143.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407.

Hermanto, N., Zuroff, D. C., Kelly, A. C., Leybman, M. J. (2016). Receiving support, giving support, and self-reassurance: A daily diary test of social mentality theory. *Personality and Individual Differences, 107,* 37-42.

Holmes, J. (2001). *The search for the secure base: Attachment theory and psychotherapy*. New York, NY: Brunner-Routledge.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Kelly, A. C., & Dupasquier, J. (2016). Social safeness mediates the relationship between recalled parental warmth and the capacity for self-compassion and receiving compassion. *Personality and Individual Differences*, *89*, 157–161.

Kelly, A. C., Zuroff, D. C., Leybman, M. J., & Gilbert, P. (2012). Social safeness, received social support, and maladjustment: Testing a tripartite model of affect regulation. *Cognitive Therapy and Research*, *36*(6), 815–826.

Kleinbaum, D. G., Kupper, L. L., Nizam, A., & Rosenberg, E. S. (2013). *Applied regression analysis and other multivariable methods*. Toronto, ON: Nelson Education.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biological Psychiatry*, *44*(12), 1248–1263.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*(3), 285–300.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Mikulincer, M., & Shaver, P. R. (2009). An attachment and behavioral systems perspective on social support. *Journal of social and personal relationships*, *26*(1), 7-19.

Muthén, L.K., & Muthén, B. O. (1998-2019). Mplus user's guide. Eighth edition. *Los Angeles, CA: Muthén & Muthén*

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition, 14*(1), 30–80.

Revelle, W. R. (2017). psych: Procedures for personality and psychological research. Version 1.6. 9. *Evanston, IL: Northwestern University*.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software. 48*(2), 1-36.

Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin, 125*(1), 3-30.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(39), 1161–1178.

Russell, J. A. (1999). The phoenix of bipolarity: A reply to Watson and Tellegen. *Psychological Bulletin, 125*(5), 611–617.

Steiger, J., H. & Lind, J. (1980, May). *Statistically-based tests for the number of common factors.* Presented at the Annual Meeting of the Psychometric Society, Iowa City, IA.

Sterba, S. K., & MacCallum, R. C. (2010). Variability in parameter estimates and model fit across repeated allocations of items to parcels. *Multivariate Behavioral Research, 45*(2), 322–358.

R Core Team (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science*, *10*(4), 297–303.

Thayer, R. E. (1997). *The origin of everyday moods: Managing energy, tension, and stress*. New York, NY: Oxford University Press.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC: American Psychological Association.

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology*, *7*, 769.

Ward, L. C. (2006). Comparison of factor structure models for the Beck Depression Inventory–II. *Psychological Assessment*, *18*(1), 81-88.

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the positive and negative affect schedule - expanded form*. Ames: The University of Iowa.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Watson, D., & Tellegen, A. (1999). Issues in dimensional structure of affect—Effects of descriptors, measurement error, and response formats: Comment on Russell and Carroll (1999). *Psychological Bulletin*, *125*(5), 601-610.

Watson, D., & Tellegen, A. (2002). Aggregation, acquiescence, and the assessment of trait affectivity. *Journal of Research in Personality*, *36*(6), 589–597.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 76*(6), 913–934.

Wood, D., Harms, P. D., Lowman, G. H., & DeSimone, J. A. (2017). Response speed and response consistency as mutually validating indicators of data quality in online samples. *Social Psychological and Personality Science, 8*(4), 454-464.

Zuroff, D. C., Sadikaj, G., Kelly, A. C., & Leybman, M. J. (2016). Conceptualizing and measuring self-criticism as both a personality trait and a personality state. *Journal of Personality Assessment, 98*(1), 14–21.

Bridge to Article Two

Article One provided strong evidence for the independence of social safeness from PA and NA factorially, in terms of total variance in social safeness explained, and in terms of safeness's unique predictions of theoretically relevant outcomes. These findings were obtained in both community and university student samples. However, social safeness is an important part of the clinical model of Compassion Focused Therapy in which it is specifically targeted as a mechanism for change in clinical symptoms (Gilbert, 2009, 2014). As such it is important (1) to assess the relationship of safeness with other affective constructs in a distressed population and (2) to assess the relationship between safeness and symptom change, which has yet to be examined by researchers. The study in Article Two was designed to address these issues.

Article Two had two aims. The first was to extend the analyses supporting the independence of social safeness from PA and NA provided in Article One in a clinically relevant sample. This sample was composed of 92 participants drawn from a mindfulness-based psychoeducational intervention, and as this sample had substantial elevations in both anxiety and depressive symptoms, this allowed us to test the assertion that safeness is distinct from PA and NA in a distressed population. In addition, to date, social safeness has never been examined in the context of a psychological intervention, and therefore it has not been established that it predicts change in clinical symptoms. The second aim of Article Two was therefore to use safeness to predict change in anxiety and depressive symptoms. This was done both with and without controlling for PA and NA in order to further distinguish social safeness from those two constructs.

Article Two

Social safeness predicts symptom change in Mindfulness Based Stress Reduction: Evidence for independence and clinical relevance

Benjamin F. Armstrong III and David C. Zuroff

McGill University

Armstrong III, B. F., Zuroff, D. C. (2019). Social safeness predicts symptom change in Mindfulness Based Stress reduction: Evidence for independence and clinical relevance. *Unpublished manuscript.*

Abstract

Social safeness is the warm soothing affect system associated with attachment processes and social connection (Gilbert et al., 2008, Gilbert, 2014). It has been the subject of a growing body of research including recent work supporting the hypothesis that it is a third affective dimension distinct from positive (PA) and negative (NA) affect. The present work aimed to further distinguish social safeness from PA and NA by showing the unique contribution of social safeness in predicting mindfulness, self-criticism, self-reassurance, and life satisfaction above and beyond PA and NA. In addition, as social safeness is a target of therapeutic change (Gilbert, 2009), its relationship to change in clinical symptoms was also examined. Intake measures of affect, personality, and clinical symptoms (depression and anxiety) were taken from a sample of 92 participants in Mindfulness Based Stress Reduction (MBSR) groups at a private clinic. Midpoint and exit measures were taken of social safeness, PA, NA, and clinical symptoms. Social safeness had significant bivariate relationships with all study variables. It uniquely predicted self-criticism, self-reassurance, and life satisfaction controlling for PA and NA. It uniquely predicted change in anxiety but not depressive symptoms over the course of MBSR programs again controlling for PA and NA.

Keywords: social safeness, positive affect, negative affect, mindfulness, self-criticism

Social safeness predicts symptom change in Mindfulness Based Stress Reduction: Evidence for independence and clinical relevance

Social safeness is the warm soothing feeling one experiences when surrounded by friends and loved ones. It is a sense that all is right with the world and that for now at least, nothing needs to be said or done. Paul Gilbert characterizes social safeness as one of three primary affective systems under his tripartite model of affect (Gilbert, 2005, 2009). These three affective systems include: positive affect (PA) that energizes us to seek out resources and opportunities (e.g., hunting for food), negative affect (NA) that empowers us to deal with threats (e.g., stalking tigers), and social safeness, Gilbert's primary contribution to this model. Social safeness is theorized to have evolved to motivate care and attachment processes first in parent-child relationships and later among members of a social group (Gilbert, 2014). As part of Compassion Focused Therapy (CFT), social safeness is a specific therapeutic target whereby clients self-activate this system by directing feelings of compassion toward the self (Gilbert 2009, 2014). Through its role in attachment security, social safeness has been shown to support exploratory motivation (Liotti & Gilbert, 2011). It has been linked to psychological well-being (Gilbert et al., 2008; Gilbert et al., 2009; Kelly & Dupasquier, 2016; Kelly, Zuroff, Leybman, & Gilbert, 2012) as well as indicators of psychopathology and psychological vulnerability (Alavi et al., 2016; Gilbert et al., 2008; Kelly et al., 2012).

Although related to social support and attachment security, social safeness is conceptually different from these constructs. It is conceptualized as an affective system, and while attachment security includes affective components like the warm feelings associated with felt security, it also includes "working models," cognitive schemas related to relationship history and beliefs (Bowlby, 1980; Gillath, Hart, Noftle, & Stockdale, 2009) that are not part of the social safeness

construct. In addition, while social safeness is related to affiliative processes including attachment (Depue & Morrone-Strupinsky, 2005; Kelly et al., 2012) it can be activated by non-attachment sources–friends, colleagues and even hierarchical relationships like those between a boss and employee (Keltner, 2003).

Social Safeness, PA, and NA

Social safeness is theorized to be distinct from PA and NA, and this has been shown in previous work (Kelly & Dupasquier, 2016; Kelly et al., 2012). It is important to note that when we discuss PA and NA in this article, we are referring to the definitions of PA and NA used by Watson, Clark, and Tellegen (1988) whereby they represent activated states or systems rather than emotional valence. Under this definition, it has been suggested that PA and NA could more properly be thought of as *Positive Activation* and *Negative Activation* as they are primarily concerned with activated states related to appetitive (for PA) or aversive (for NA) stimuli. Social safeness is positively related to PA and negatively related to NA, and these relationships have important unique components whereby social safeness predicts each controlling for the other (Armstrong, Nitschke, Bilash, & Zuroff, 2019). Further, Armstrong et al. (2019) found that PA and NA together accounted for only 35% to 61% of the variance in social safeness, depending on the sample, suggesting independence. Based on the relationships between safeness and PA and NA, it has been suggested that social safeness acts as a buffer against stress and encourages approach motivation (Armstrong et al., 2019), which is in line with the previous theoretical assertion that social safeness helps to create a secure base that encourages exploratory approach behavior (Gilbert, 2014). These relationships have been shown at both the between-persons and within-person levels (Armstrong et al., 2019; Kelly et al., 2012). Within-person, daily measures of social safeness have been associated with both higher PA and lower NA. That social safeness

would be associated with lower NA is intuitive, but the association with higher PA is potentially inconsistent with the original theory of Gilbert et al. (2008) which posited that safeness would de-activate this motivational system in-the-moment. It should be noted that both studies reporting within-person relationships between social safeness and PA measured at the day-level. In-the-moment, they may relate differently.

Social Safeness, Life Satisfaction, and Mindfulness

It has been asserted that the safeness/soothing system is related to the expansive feelings associated with mindfulness and meditative practices (Gilbert, 2014). Mindfulness has a variety of definitions that include different possible aspects of the construct. The definition we use (Lau et al., 2006) includes two independent dimensions. *Attention control* is one's ability to hold attention on the object or target of one's choice. This dimension is sometimes called *awareness* or *curious awareness* (Cardaciotto, Herbert, Forman, Moitra, & Farrow, 2008; Lau et al., 2006). *Decentered awareness* is one's ability to maintain a relaxed accepting attitude toward one's experience (Lau et al., 2006) rather than judging or comparing, which are associated with threat (Gilbert, 2014). This dimension is variously called acceptance, non-judgment, and/or non-reactivity (Cardaciotto et al., 2008; Lau et al., 2006). Currently only one study, Martins et al. (2019), has directly linked social safeness to mindfulness, finding a strong positive association with decentered awareness (attention control was not measured). Other recent work has also linked decentered awareness to *social connectedness*, a construct that is conceptually similar to social safeness (Adair, Fredrickson, Castro-Schilo, Kim, & Sidberry, 2018). There is currently no work reporting on the relationship between safeness and attention control.

A recent line of research (A. Akin & Akin, 2015; U. Akin & Akin, 2016; Akın & Akın, 2016) investigated the link between social safeness and life satisfaction. Social safeness and life

satisfaction showed strong associations in each study, and safeness mediated the relationship of life satisfaction with a variety of behaviors including Facebook use and vengeance-seeking. Social safeness may act as an affective mechanism by which life-satisfaction, a personality trait, affects momentary behaviors

Self-criticism, Self-reassurance, and Compassion Focused Therapy

The tripartite model of affect including the social safeness system is explicitly taught to clients in Compassion-Focused Therapy (CFT), a system of therapy that emphasizes compassion for the self and others (Gilbert, 2009; Gilbert, 2014). As part of this therapy, one is taught to treat oneself with *self-compassion* or *self-reassurance*, which is the ability to be self-validating, supportive, and compassionate with the self (Gilbert, 2004; Gilbert 2014). Theoretically, this will allow one to activate one's own social safeness system (Gilbert, 2014), and this is supported by strong associations between social safeness and measures of both self-compassion and self-reassurance (Gilbert, 2008; Kelly & Dupasquier, 2016; Mendez, Ferreira & Trindade, 2018).

Treating oneself with compassion or reassurance is seen in CFT as a way of dealing with self-criticism, which is viewed as a protection against perceived threat (Gilbert, 2014). By viewing the pain behind these perceived threats compassionately, one is able to "[recode] the emotional memory with the new affect processing […] of the affiliative system that was evolved to down regulate threat" (Gilbert, 2014, p.32). Through this reprocessing, social safeness is theorized to reduce self-criticism (by reducing threat), and this is supported by its medium to large negative associations with self-criticism (Gilbert, 2008; Kelly et al., 2012; Richter, Gilbert & McEwan, 2009).

Social safeness and symptoms of anxiety and depression.

Social safeness has clinical relevance through its proposed role as a soothing system that is hypothesized to reduce anxiety and distress by inhibiting the NA system. Thus, one would expect that social safeness would have a strong negative relationship with anxiety, and this is supported by previous research showing a strong negative association between safeness and anxiety (Gilbert, 2010; Gilbert et al., 2008; Kelly et al., 2012). Safeness's theoretical links with depressive symptoms are less direct. Through its association with attachment, social safeness is theorized to support the cultivation of a *secure base* that encourages exploration as well as "drive functions, facilitating behaviours for seeking and acquiring skills and resources and facing challenges" (Gilbert, 2014, p. 19). Thus, although Gilbert asserts that social safeness should de-activate approach motivation (PA) in the moment, it should support PA on longer timeframes. As low PA is theoretically and empirically linked to depression and depressive symptoms (Clark & Watson, 1991), one would expect social safeness to have a negative relationship with depressive symptoms. Again, this is supported by previous research showing a medium-sized negative association between social safeness and depressive symptoms (Gilbert, 2010; Gilbert et al., 2008; Kelly et al., 2012). In addition, self-compassion and the associated feelings of social safeness are also linked to a reduction in self-critical thoughts that are strongly associated with depression (Gilbert, 2014). Social safeness has been associated with reduced fluctuations in mood, including depressive symptoms, in bipolar disorder (Gilbert et al., 2009).

The Present Study

In this study, we sought to investigate the clinical relevance of social safeness both in its own right and as compared to the affective dimensions PA and NA. This work was conducted in the context of Mindfulness Based Stress Reduction (MBSR) groups in which participants were asked to fill out surveys related to their mood and personality before, at the midpoint, and after

their program. MBSR is a 7-week psychoeducational program created by Jon Kabat-Zinn (Kabat-Zinn, 1982) that gives participants a broad overview of mindfulness as well as specific instruction in several types of guided and unguided meditation including: body awareness, seated meditation, walking meditation, and mindful yoga. While MBSR is generally considered a psychoeducation program rather than a form of psychotherapy, the sample exhibited elevations in depressive and anxiety symptoms.

We began by examining the links between social safeness and clinically relevant personality constructs at intake and then investigated safeness's ability to predict change in anxiety and depressive symptoms over the course of an intervention controlling for PA and NA. It is important to control for PA and NA in this context, as two-dimensional models of affect would predict that the relationship of any affect to an outcome would be explained by its relationships with PA and NA (Russell, 1980; Tellegen, 1985). This was the first time to our knowledge that social safeness has been used to predict intake levels of, as well as treatment-related changes in, depressive and anxiety symptoms while controlling for PA and NA.

The relationship of social safeness was examined at intake with: self-criticism and self-reassurance, mindful attention and acceptance, and life satisfaction. As mentioned above, social safeness has strong theoretical and empirical links to self-criticism and self-reassurance. Thus we hypothesized that these relationships would replicate here and hold when controlling for PA and NA. Likewise, social safeness has empirical associations with mindful acceptance. As mindful acceptance is the conceptual opposite of judgment and criticism (Cardaciotto et al., 2008; Lau et al. 2006), which are associated with threat (Gilbert, 2014), we hypothesized that social safeness would have a strong positive relationship with mindful acceptance and that this relationship would hold controlling for PA and NA. Social safeness has also been shown to have a strong

positive association with life satisfaction (A. Akin & Akin, 2015; U. Akin & Akin, 2016; Akın & Akın, 2016) as well as a theoretical association through its links to attachment security which itself is associated with life satisfaction (Lavy & Littman-Ovadia, 2011). Thus, we hypothesized that social safeness would have a strong positive association with life satisfaction and that this would hold when controlling for PA and NA. We had no clear hypothesis regarding mindful attention as it has not been examined by prior work, and it is not clear how social safeness would relate to the ability to focus attention.

Because social safeness is theorized to drive therapeutic processes as part of CFT, we also examined social safeness's ability to predict change in anxiety and depressive symptoms. This was done by using safeness at midpoint to predict exit measures of depression and anxiety symptoms controlling for intake measures of those symptoms. The midpoint measure was used (1) to reflect the temporal ordering of the presumed causal sequence from group participation to safeness to symptom change and (2) to address the potential confounding of safeness with exit levels of symptoms. In this way, we were able to assess the extent to which the level of safety a person had established by the program midpoint was related to changes in their anxiety and depressive symptomology. A limitation of this approach is that the role of social safeness may be confounded with early symptom change (see: Feely, DeRubeis & Gelfand, 1999).

Method

Participants were 92 English-speaking North American adults recruited in-person from Mindfulness Based Stress Reduction (MBSR) groups offered at a private psychology clinic in Montreal. Participants ranged in age from 19 to 69 (M = 40.76, SD = 13.21). They were primarily female (N = 65), and the majority had completed at least some post-secondary education (93%). See Table 1 for additional demographic information.

Not all participants that started the study completed their participation. This resulted in a reduced participant pool at midpoint of 67 participants and at exit of 65 participants. At exit, participants ranged in age from 18 to 69 (M = 40.76, SD = 13.21) and were primarily female (N = *48*), with the majority completing some post-secondary education (92%). Demographic information for this reduced sample is also available in Table 1.

The sample had elevated levels of depressive symptomology with the sample mean (M = 6.47, SD = 4.33) exceeding the cutoff (5) for mild depression on the Patient Health Questionnaire (PHQ9: Kroenke, Spitzer, & Williams, 2001). The national average in the United States is 3.1 for women and 2.7 for men (Kocalevent, Hinz, & Brähler, 2013). In this sample, 15 participants (16%) met the cutoff for major depression (15) on an instrument with 88% sensitivity and 88% specificity for that cutoff (Kroenke et al., 2001). The sample had elevated levels of anxious symptomology with the sample mean (M = 7.64, SD = 4.42) exceeding the cutoff (5) for mild generalized anxiety on the Generalized Anxiety Disorder scale (GAD7: Spitzer, Kroenke, Williams, & Löwe, 2006). The national average for the United states is 3.2 for women and 2.7 for men (Löwe et al., 2008). In this sample, 24 participants (26%) met the cutoff for severe anxiety (15) on an instrument with 89% specificity and 82% sensitivity at that cutoff (Spitzer et al., 2006).

Procedure

Participants were recruited from MBSR groups in collaboration with group leaders at a private psychotherapy clinic in Montreal. Recruitment involved an in-person presentation of the study during an information session, held one week before the start of the program. This presentation was delivered while the group leader was not present, and participants were assured that their participation and responses would be kept confidential and not shared with anyone

including their group leader. Interested participants entered their name, email and phone number on a sign-up form. The next day, those who had given their contact information were contacted by email with instructions for how to register on the study web site. The study was conducted entirely online and consisted of intake, midpoint and exit measures of affective, clinical, and personality constructs. Participants were given some daily and weekly measures that are not relevant to the present work. Participants received a reminder by email or text (their choice) each day at 5pm alerting them to complete their present study measures. Participants were compensated incrementally based on the amount of the study completed with up to $56 possible for completing the intake ($10), midpoint ($5), exit ($5), eight session ($1), and 50 daily questionnaires ($.50 + $3 for 90% completion). The study protocol was approved by the research ethics board at McGill University (REB File #: 152-0816).

Groups

Participants were recruited from 18 groups between Fall, 2016 and Fall, 2018. Each group consisted of 10-20 participants, and recruitment typically resulted in 4-8 signups, although as few as one or as many as 10 occurred. Groups were conducted in either English or French. Participation by non-native English speakers was permitted if they endorsed that they were comfortable with the English-language study materials.

Group leaders were certified phase one (or higher) instructors in MBSR according to the Mindfulness Based Professional Training Institute. In total, nine instructors delivered MBSR courses over the course of the study. Seven of the instructors hold doctoral degrees in psychology. In some of the courses there was a doctoral student assisting with the group for training purposes supervised by a certified instructor.

Measures

Reliabilities for all measures were computed using McDonald's Omega (McDonald, 1999) calculated in the statistical program *R* (R Core Team, 2017) with the package *psych* (Revelle, 2018). Omega has been shown to be particularly well-suited to structural equation models, like those used in the present work (Zinbarg, Revelle, Yovel, & Li, 2005). Only the intake, midpoint, and exit measures germane to the present work are listed below.

Positive and negative affect. Positive and negative affect were assessed using the Positive And Negative Affect Schedule (PANAS: Watson et al., 1988). The PANAS is a 20-item measure of two affective dimension: positive and negative affect. Positive affect is the extent to which one is engaged and enthusiastic. Negative affect is the extent to which one is distressed and/or upset. Each subscale consists of 10 mood adjectives (e.g., *enthusiastic*, *nervous*). Participants rated the extent to which they were experiencing each mood adjective using a 5-point Likert scale ranging from 1 (*very slightly or not at all*) to 5 (*extremely*) for the time period in *the past month.* McDonald's omega was .90 for positive affect items and .85 for negative affect items at intake.

Social safeness. Social safeness was assessed using the Types of Positive Affect Scale (TPAS: Gilbert et al., 2008). The TPAS is an 18-item measure that assesses three types of positively valenced affect: active, relaxed and safe. Although not as widely used as the Social Safeness and Pleasure Scale (SSPS: Gilbert et al., 2009), the safeness items have shown good reliability and factorial validity in previous work (Armstrong et al., 2019). Only the safeness items, which measure the extent to which one feels warm and secure were used in these analyses. The scale is comprised of four mood adjectives: *safe*, *secure*, *warm*, and *content*. Participants rated the extent to which they were experiencing each mood adjective using a 5-point Likert

scale ranging from 0 (*Not at all characteristic of me*) to 4 (*Very characteristic of me*) for the time-frame *in general*. McDonald's omega for social safeness items was .86 at intake.

Mindfulness. Trait mindfulness was assessed using the Philadelphia Mindfulness Scale (PHLMS: Cardaciotto et al., 2008). The PHLMS is a 20-item measure that assesses two aspects of mindfulness: attention control and decentered awareness. Subscales consist of statements describing experience related to attention (e.g., *I am aware of what thoughts are passing through my mind.*) and acceptance (e.g, *There are aspects of myself that I don't want to think about.*). The PHLMS has received positive conceptual comparisons to other mindfulness instruments (Baer, 2011; Baer, Walsh, & Lykins, 2009). It has shown a clear two-factor structure and good reliability in multiple languages (Tejedor, Feliu-Soler, & Pascual, 2014; Teixeira, Ferreira, & Pereira, 2017). It should be noted that acceptance items in this scale are worded to reflect a lack of acceptance rather than the presence of acceptance. To avoid confusion, all scores were reversed so higher acceptance reflects higher rather than lower acceptance. This is consistent with what was reported in the original scale validation (Cardaciotto et al., 2008). Participants rated how well the statements describe their experience using a 5-item Likert scale ranging from 1 (*Never*) to 5 (*Very Often*). McDonald's omega was .87 for attention control items and .88 for decentered awareness items at intake.

Depressive symptoms. Severity of depressive symptoms was assessed using the Patient Health Questionnaire 9-item version (PHQ9: Kroenke et al., 2001). The PHQ9 asks the extent to which one has been bothered by problems that closely mirror DSM-V (American Psychiatric Association, 2013) criteria for depression (e.g., *Little interest or pleasure in doing things*). Its validity, sensitivity, and specificity have been extensively examined, and it has been found to be a valid and reliable measure of depression in the general population (Kocalevent, Hinz, &

Brähler, 2013). Items are rated on a 4-point Likert scale from 0 (*Not at all*) to 3 (*Nearly every day*) for the time-frame *over the past two weeks*. McDonald's omega was .79 at intake.

Anxiety symptoms. Anxiety symptoms were assessed using the Generalized Anxiety Disorder 7-item version (GAD7: Spitzer et al., 2006). The GAD7 asks the extent to which one has been bothered by problems that closely mirror DSM-V (American Psychiatric Association, 2013) criteria for generalized anxiety (e.g., *Feeling nervous, anxious, or on edge*). Its validity, sensitivity, and specificity have been extensively examined, and it has been found to be a valid and reliable measure of anxiety in the general population (Löwe et al., 2008). Items are rated on a 4-point Likert scale from 0 (*Not at all*) to 3 (*Nearly every day*) for the time-frame *over the past two weeks*. McDonald's omega for anxiety was .86 at intake.

Self-criticism and Self-reassurance. Self-reassurance was assessed using the Forms of Self-Criticizing and Self-Reassuring Scale (FSCSRS: Baião, Gilbert, McEwan, & Carvalho, 2014). The FSCSRS is a 22-item measure that assesses three aspects related to self-criticism: inadequate self, hated self, and self-reassurance. Inadequate self is a "sense of feeling internally put-down and rendered inadequate by failures and setbacks" (Gilbert, 2004, p.38). Hated self is a "destructive, disgust-based response to setbacks (rather than inadequate) characterized by self-dislike and an aggressive/sadistic/persecuting desire to hurt the self" (Gilbert, 2004, p.38). These two subscales correlate highly and recent work has supported merging them into a valid single *self-criticism* subscale (Halamová et al., 2018). *Self-reassurance* is one's ability to be compassionate and supportive with the self in the face of struggle. Items in the FSCSRS consist of statements about how one treats oneself (e.g., Self-criticism: *I am easily disappointed in myself;* Self-reassurance: *I am able to care and look after myself*). Items are rated for the time-frame *in general* on a 5-point Likert scale ranging from 0 (*Not at all like me*) to 4 (*Extremely like*

me). McDonald's omega for self-reassurance items was .85 at intake. Omega for self-criticism items was .89 at intake. Self-criticism and self-reassurance scores were strongly correlated at intake, r (86) = -.65, $p < .001$.

Life Satisfaction. Life satisfaction was assessed using the Satisfaction with Life Scale (SWLS: Diener, Emmons, Larsen, & Griffin, 1985). The SWLS is a 5-item measure that assesses the extent to which a person is satisfied with their life. Its factor structure and validity have been extensively examined and found to be consistently good in a wide variety of populations (Pavot & Diener, 2009; Shevlin, Brunsden, & Miles, 1998). Items consist of statements about how one views one's life ("I am satisfied with my life.*"*). Items are rated *in general* on a 7-point Likert scale ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*). McDonald's omega for life satisfaction was .88 at intake.

Data Analytic Plan

To establish the validity of our measurement model, confirmatory factor analyses (CFA) were conducted using the statistical program *R* (R Core Team, 2017) and the package *lavaan* (Rosseel, 2012). Subsequent analyses were conducted in SEM using this measurement model or its parts (e.g., social safeness only) in predicting observed personality variables or observed clinical outcomes.

The measurement model for this CFA used an item parceling strategy for PA and NA employed in previous work (Armstrong et al., 2019; Dunkley et al., 2017; Dunkley, Ma, Lee, Preacher, & Zuroff, 2014). Parceling offers a number of psychometric benefits including improved indicator reliability and the creation of more properly continuous indicators suitable to maximum likelihood estimation (Little, Rhemtulla, Gibson, & Schoemann, 2013). The parceling procedure was to select every third item from each subscale. This led to two three-item parcels

and one four-item parcel for each PANAS subscale. Prior work has suggested that for a well-established unidimensional factor structure like that of the PANAS subscales, this is unlikely to change the results of subsequent analyses (Dunkley et al., 2017; Sterba & MacCallum, 2010). We did not employ a parceling structure for social safeness as it had only four items.

Comparative fit index (CFI; Bentler, 1990) and root mean-square error of approximation (RMSEA; Steiger & Lind, 1980) are reported for measurement models. Recent recommendations for model-fit cutoff criteria are much stricter than previous guidance (Hu & Bentler, 1999). Thus, we will report fit as *excellent* if it meets stricter criteria (CFI $\geq$ .95, RMSEA $\leq$ .06), *good* if it meets previous more liberal criteria (CFI $\geq$ .90, RMSEA $\leq$ 08), *adequate* if it meets one criterion and just misses the other, and *marginal* if it comes close but does not meet either criterion. This is the same interpretation strategy used in Armstrong et al., (2019) with the addition of the *adequate* category. Correlations and standardized covariances are described according to the recommendations of Cohen (1992, 2013) with .10 $\leq r <$ .30 considered *small*, .30 $\leq r <$ .50 considered *medium*, and *r*s $\geq$ .50 considered *large*.

After specifying measurement models, bivariate and unique relationships among latent social safeness, PA, and NA are reported. Subsequently, SEM was used to investigate the bivariate and unique relationships of social safeness with personality constructs. In these models and those that follow, observed variables were used for all personality and symptom constructs rather than latent variables. This was done because it is not recommended to create latent variables with many indicators particularly when sample size is relatively small (Kline, 2015), and item heterogeneity did not permit parceling (Little et al., 2013). For each of these constructs, two models were conducted: one model with latent social safeness predicting the observed construct score to investigate the bivariate relationship and a second model with latent social

safeness predicting the observed construct controlling for latent PA and NA to investigate safeness's unique relationship.

Finally, midpoint latent social safeness was used to predict each observed symptom measure at exit (i.e., depression or anxiety) controlling for the intake assessment of that symptom measure. This analysis was first conducted with social safeness as the sole predictor and then repeated controlling for latent PA and NA at the program midpoint.

Results

Measurement models consisting of latent PA, NA and social safeness were evaluated using MLR estimation to establish the relationships among these constructs at intake and then again at the program midpoint. The model produced good to good fit at intake, $x^2(45) = 431.37$, CFI = .96, RMSEA = .07, and marginal to poor fit at the program midpoint, $x^2(45) = 394.55$, CFI = .83, RMSEA = .17, though the addition of a single covariance in this midpoint model between safeness items one (secure) and two (safe) improved the fit to good, $x^2(45) = 394.55$, CFI = .97, RMSEA = .08. Subsequent analyses using program midpoint social safeness alongside PA and NA included this covariance. All standardized factor loadings were above .70.

Associations among PA, NA, and Social Safeness

Bivariate associations. The bivariate relationships among PA, NA, and social safeness were calculated using both observed correlations of composite scores and standardized covariances (correlations) of latent variables at intake (see Table 3). Latent variable associations were highly similar to observed correlations, and the pattern of relationships among the three variables across intake and the program midpoint was consistent. Consistent with previous work, PA and NA showed a small negative association. Latent PA accounted for 51% of the variance in latent social safeness and latent NA accounted for 34% of the variance in latent safeness. These

findings were taken as evidence that social safeness is related to but not redundant with PA or NA individually.

Partial regression coefficients. The unique associations of PA and NA in predicting social safeness were calculated by regressing latent safeness simultaneously on latent PA and NA in SEM. It was found that PA had a large and significant positive unique association with social safeness, $\beta = .59$, $z = 4.58$, $p < .001$, and NA had a unique negative one, $\beta = -.32$, $z = -2.58$, $p = .010$. PA and NA showed a small and negative significant association in this model, $r = -.28$, $z = -2.22$, $p = .026$. PA and NA together accounted for 56% of the variance in social safeness. The unique relationships of safeness with PA and NA combined with the substantial percentage of variance in social safeness not accounted for by PA and NA replicated previous work showing that social safeness is not fully predicted by PA and/or NA.

Personality Predictors of Social Safeness

Both bivariate and unique relationships (controlling for both PA and NA) were calculated for social safeness with several personality variables at Time 1 including: self-criticism, self-reassurance, life satisfaction, and mindfulness. These relationships are summarized in Table 4. As expected, social safeness showed a large negative bivariate association with self-criticism and a large positive bivariate association with self-reassurance. Safeness continued to show medium unique relationships with both variables when controlling for PA and NA, indicating that safeness is independently related to the way one treats oneself.

Social safeness showed a medium bivariate association with mindful attention, but this association was no longer significant when controlling for PA and NA. Safeness showed no relationship with mindful acceptance. Any relationship between safeness and mindfulness as measured in this study appears to be explained by the variance it shares with PA and/or NA.

Social safeness also showed a large positive association with life satisfaction. This association remained large when controlling for PA and NA, indicating an important unique relationship with well-being more generally.

Social Safeness Predicting Symptom Change

Intervention check. To assess whether change was achieved over the course of the MBSR program, descriptive statistics and mean differences were calculated for each affect and clinical outcome variable. Paired samples t-tests were conducted to assess the significance of the mean differences. These statistics are available in Table 2. All study variables showed significant change from intake to exit. Positively valenced measures–social safeness and PA–increased over the course of the program, and negatively valenced measures–anxiety, depressive symptoms and NA–decreased. Mean levels in both anxiety and depressive symptoms decreased to below the threshold for mild symptomology (5). However, both were still above population averages for those measures.

The bivariate and unique relationships of social safeness with depressive and anxiety symptoms were calculated at intake. These analyses were conducted in SEM with latent PA, NA and safeness correlated with observed depressive and anxiety symptoms. These relationships are summarized in Table 5. Social safeness had large inverse relationships with both anxiety and depressive symptoms. PA and NA both had medium to large relationships in the expected directions with these symptoms. When PA and NA were controlled, social safeness showed no significant unique relationship with either depressive or anxiety symptoms indicating that the relationships between social safeness and both sets of symptoms at intake were accounted for by safeness's shared variance with PA and/or NA.

In order to assess the ability of social safeness to predict symptom change, social safeness at the program midpoint was used to predict first depressive and then anxiety symptoms at exit controlling for the intake symptom measure. This procedure was used to provide evidence that the degree of social safeness established by the program midpoint was predictive of symptom change. Analyses were conducted first with only social safeness and then controlling for PA and NA at the program midpoint giving both the individual and unique contribution of safeness in predicting change in depressive and anxiety symptoms over the course of the program. These relationships are summarized in Table 6. Social safeness at the midpoint predicted greater reduction in both anxiety and depressive symptoms from intake to exit. When controlling for PA and NA at the midpoint, social safeness no longer predicted a reduction in depressive symptoms. It continued to predict a reduction in anxiety symptoms and the effect remained large. These findings suggest that the level of safeness achieved by the midpoint in treatment may be an important unique predictor of at least one clinically relevant outcome, reduction in anxiety symptoms.

Discussion

The aims of this study were (1) to assess the bivariate relationships of social safeness with a variety of clinically relevant personality and symptomology constructs, including change in depression and anxiety symptoms over the course of an MBSR intervention and (2) to assess the *unique* relationship of social safeness with these same constructs above and beyond PA and NA. First, the bivariate and unique relationships of social safeness with personality constructs were examined at intake. Next, social safeness was used to predict change in symptoms over the course of the program both individually and controlling for PA and NA.

In keeping with previous work (Armstrong et al., 2019; Kelly et al., 2012), it was found that social safeness was not redundant with PA and NA, the two constructs combined accounted for approximately half of the reliable (non-error) variance in social safeness. Given previous work showing that social safeness contributes important unique variance in predicting a variety of outcome measures (Armstrong et al., 2019), it appears reasonable to conclude that social safeness is indeed an affective system distinct from PA and NA, as suggested by Gilbert et al. (2008). If social safeness is an independent affective system with important links to attachment and mental health both theoretically (Gilbert, 2014; Gilbert et al., 2008) and empirically (Armstrong et al., 2019; Gilbert et al., 2009; Kelly & Dupasquier, 2016; Kelly et al., 2012), it becomes important to assess its relationship with a wider range of clinically relevant constructs including outcome measures like depressive symptoms and anxiety.

Toward this end, we examined the relationship of social safeness with self-criticism and self-reassurance. These constructs show strong relationships with various markers of psychopathology (Baião et al., 2014) and are theoretically important targets of therapeutic change in CFT. It was found that safeness had a strong negative relationship with self-criticism and a strong positive relationship with self-reassurance, and these relationships held when controlling for PA and NA. Although we cannot determine directionality from this design, self-criticism is theorized to undermine one's ability to self-soothe and self-generate feelings of safety and security while self-reassurance is a process that creates those feelings (Baião et al., 2014; Gilbert, 2009). In CFT, compassionate self-reassurance is used to activate the social safeness system for the purposes of self-soothing and to reduce self-criticism. That social safeness showed the theorized links with self-reassurance and self-criticism is important

evidence in support of this model. That these relationships held when controlling for PA and NA is important additional evidence that social safeness is an independent affective dimension.

Next we examined the relationship of social safeness with measures of mindful awareness, mindful acceptance and life satisfaction. Social safeness had an unpredicted, medium relationship with mindful attention, and it failed to show the predicted relationship with mindfulness acceptance. The lack of a relationship with acceptance may have been due to a lack of statistical power as our sample was somewhat small. It could also have been related to the approach of the PHLMS, which measures acceptance by its absence. The significant association between social safeness and mindful acceptance found previously (Martins et al., 2019) used an instrument that measures acceptance by its presence (TMS: Lau et al., 2006). That the relationship with mindful attention was wholly explained by the variance social safeness shares with PA and NA may indicate that these attention processes are primarily driven by the appetitive processes associated with PA. In fact attentiveness has been cited as an individual emotion subsumed by PA (Watson & Clark, 1999), and the subscale includes the items *attentive* and *alert*, which are both relevant to attention control. Mindful awareness may be driven by the motivated desire to focus one's attention rather than the presence of social connectedness that characterizes social safeness. Social safeness had a strong relationship with life satisfaction that was unchanged when controlling for PA and NA. The sense that life is satisfying and complete seems to be closely tied to social safeness's sense of connectedness and interpersonal warmth. This implies that there is a social or connectedness element to life satisfaction that is very much in keeping with the idea that a sense of belonging and/or relatedness to others is essential for overall well-being (Baumeister & Leary, 1995; Ryan & Deci, 2000).

Last, we examined the relationship of social safeness with clinically relevant outcomes: reductions in depressive and anxiety symptoms. It should be noted that while this was not a psychiatric sample comprised of individuals with diagnosed disorders, it did include a substantial number of participants whose levels of anxiety and depressive symptoms exceeded suggested cut-offs for the likely presence of generalized anxiety disorder or major depression. Anxiety symptoms, depressive symptoms, and all three affective constructs showed significant change in the expected direction between intake and exit. Social safeness showed strong relationships with change in both sets of symptoms, though the relationship held only for anxiety and not depressive symptoms when controlling for PA and NA. Theoretically, this may be explained by the assertion that social safeness supports reductions in the activity of other affect systems (e.g., anxiety produced by the NA system) in the moment and increased motivation and drive, associated with reduced depressive symptoms, only on longer timescales (Gilbert, 2014). It is possible that the seven-week MBSR program was not long enough for social safeness to show a relationship with depressive symptoms above and beyond PA and NA. It is also possible that social safeness does not have an independent effect on depressive symptoms above and beyond PA and NA in general or in this sample. Future work might use experimental designs to look at the causal order of these variables in order to further disentangle the links between safeness and clinical symptomology.

Also counter to our expectations, social safeness did not show the expected unique relationships with anxiety or depressive symptoms at intake (although it did show bivariate relationships). This brings up an important theoretical and analytical issue regarding the relationships of PA and NA with clinical symptoms. Anxiety is a major component of the NA construct, with NA essentially being the superset of psychological distress that includes anxiety.

For depressive symptoms, low PA represents one form of depressive symptomology: the anhedonic type associated with low motivation (Watson et al., 1991). Thus, when we control for PA and NA in predicting depressive symptoms and anxiety, we may in fact be predicting our outcomes controlling for other measures of those same outcomes. Consequently, the analyses predicting those outcomes with social safeness controlling for both NA and PA may be highly conservative, and the fact that social safeness predicted change in anxiety with those controls may be an even stronger indicator of its independence as a system and its clinical relevance. The bivariate relationships between social safeness and clinical outcomes may also be worth considering. Social safeness did indeed predict change in both anxiety and depression. The relationship was small for depression and large for anxiety. Combined with previous measurements of the relationship between social safeness and these constructs (Gilbert et al., 2008; Kelly et al., 2012), and with previous research showing that it is associated with reduced distress and increased engagement (Armstrong et al., 2019), these links with anxiety and depression indicate that social safeness is an important predictor of clinical change.

In general, the unique relationships of social safeness with several clinically relevant personality variables including self-criticism and life satisfaction make a compelling case that safeness is both (1) distinct from PA and NA and (2) clinically relevant. This was further supported by its unique relationship with change in anxiety over the course of the intervention. As the target of CFT interventions, social safeness shows all of the characteristics that theory would predict. If we allow that it is distinct from PA and NA and go on to look at its individual relationships with our study variables, we see that it is an important indicator for a wide range of personality and symptomology constructs. Taking this alongside its theoretical relationship with attachment processes which are theorized to be an important part of the development of the

therapeutic relationship (Mallinckrodt, 2010; Mallinckrodt & Jeong, 2014), social safeness begins to look like an important indicator of both therapeutic outcome and process and a viable candidate for any standard battery of therapy assessments.

Limitations

Important limitations to the present work include the modest sample size and aspects of the study design. Attrition reduced the sample to approximately 65 individuals by the study's end. This limited our ability to use latent variable models for our personality variables and outcomes. As a result, the true relationships between social safeness and these variables may be larger than those we report, as latent variable modeling can reduce error and disattenuate associations. The sample size also reduced study power making it difficult to detect smaller effect sizes. As a result, *r*s smaller than approximately .25 did not reach statistical significance in our models causing us to be unable to reject the null hypothesis of no relationship though one might have been present in a larger sample from the same population.

Aspects of the study design and population limit what we can say about symptom change in other populations. Although the data did come from a sample with elevated levels of depressive symptoms and anxiety, no diagnostic interview was administered to determine whether they would meet criteria for a psychiatric disorder. In addition, our predominantly affluent, educated, and female sample interested in learning mindfulness is by no means representative of the population at large. Moreover, only a percentage of each group agreed to participate which could have led to some self-selection bias in the findings. The results we found may also be specific to mindfulness interventions in particular rather than therapeutic interventions more generally. The study was also not a randomized control, meaning that at least some portion of the change in anxiety that was predicted by social safeness could have been

related to factors other than group participation. Lastly, it is possible that our measure of social safeness at the program midpoint was to some degree confounded with early symptom change, a problem with midpoint measurement outlined by Feeley et al., (1999).

Conclusion and Future Directions

In a sample of MBSR participants with clinical elevations in depressive and anxiety symptoms, it was found that social safeness had unique relationships with self-criticism, self-reassurance, and life satisfaction controlling for PA and NA. Social safeness measured at the midpoint of the program predicted depressive and anxiety symptoms at exit controlling for intake measures of those symptoms, indicating that social safeness has a relationship with symptom change. These change findings held for anxiety but not depressive symptoms when controlling for PA and NA. Taken together these findings provide further evidence that social safeness is distinct from PA and NA and does not fit neatly into that two-dimensional affect model. The idea that social safeness is an important predictor of clinical change was also supported. There is now growing support for the clinical importance of social safeness. As the construct can be measured with as few as four items, clinical and attachment researchers would do well to consider including a measure of social safeness in their research designs.

References

Adair, K. C., Fredrickson, B. L., Castro-Schilo, L., Kim, S., & Sidberry, S. (2018). Present with you: Does cultivated mindfulness predict greater social connection through gains in decentering and reductions in negative emotions? *Mindfulness, 9*(3), 737–749.

Akin, A., & Akin, U. (2015). The mediating role of social safeness on the relationship between Facebook® Use and life satisfaction. *Psychological Reports*, *117*(2), 341–353.

Akin, U., & Akin, A. (2016). Investigating the mediator role of social safeness on the relationship between forgiveness and life satisfaction. *Anales de Psicología / Annals of Psychology*, *32*(2), 528–534.

Akın, U., & Akın, A. (2016). Examining mediator role of the social safeness on the relationship between vengeance and life satisfaction. *Social Indicators Research*, *125*(3), 1053–1063.

Alavi, K., Moghadam, A., Ali, M., Rahiminezhad, A., Farahani, H., & Modares Gharavi, M. (2016). Psychometric properties of Social Safeness and Pleasure Scale (SSPS). *Journal of Fundamentals of Mental Health*, *19*(1), 5–13.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Armstrong, B. F., Nitschke, J. P., Bilash, U., & Zuroff, D. C. (under review). An affect in its own right: Investigating the relationship of social safeness with positive and negative affect. *Personality and Individual Differences.*

Baião, R., Gilbert, P., McEwan, K., & Carvalho, S. (2014). Forms of self-criticising/attacking & self-reassuring scale: Psychometric properties and normative study. *Psychology and Psychotherapy: Theory, Research and Practice*, *88*(4), 438-452.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529.

Baer, R. A. (2011). Measuring mindfulness. *Contemporary Buddhism*, *12*(01), 241–261.

Baer, R. A., Walsh, E., & Lykins, E. L. (2009). Assessment of mindfulness. In F. DiDonna (Ed.), *Clinical handbook of mindfulness* (pp. 153–168). New York, NY: Springer.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238.

Bergomi, C., Tschacher, W., & Kupper, Z. (2013). The Assessment of mindfulness with self-report measures: Existing scales and open issues. *Mindfulness*, *4*(3), 191–202.

Bowlby, J. (1980). *Attachment and loss*. New York, NY US: Basic Books.

Cardaciotto, L., Herbert, J. D., Forman, E. M., Moitra, E., & Farrow, V. (2008). The assessment of present-moment awareness and acceptance: The Philadelphia Mindfulness scale. *Assessment*, *15*(2), 204–223.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*(3), 316.

Depue, R. A., & Morrone-Strupinsky, J. V. (2005). A neurobehavioral model of affiliative bonding: Implications for conceptualizing a human trait of affiliation. *Behavioral and Brain Sciences*, *28*(3), 313–350.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*(1), 71–75.

Dunkley, D. M., Lewkowski, M., Lee, I. A., Preacher, K. J., Zuroff, D. C., Berg, J.-L., … Westreich, R. (2017). Daily stress, coping, and negative and positive affect in depression: Complex trigger and maintenance patterns. *Behavior Therapy*, *48*(3), 349–365.

Dunkley, D. M., Ma, D., Lee, I. A., Preacher, K. J., & Zuroff, D. C. (2014). Advancing complex explanatory conceptualizations of daily negative and positive affect: Trigger and maintenance coping action patterns. *Journal of Counseling Psychology, 61*(1), 93.

Feeley, M., DeRubeis, R. J., & Gelfand, L. A. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 67,* 578 –582.

Gilbert, P. (2005). Compassion and cruelty: A biopsychosocial approach. In P. Gilbert (Ed.), *Compassion: Conceptualisations, research, and use in psychotherapy,* (pp. 21–86). London: Routledge.

Gilbert, P. (2009). Introducing compassion-focused therapy. *Advances in Psychiatric Treatment*, *15*(3), 199–208.

Gilbert, P. (2010). *Compassion Focused Therapy: Distinctive features.* New York, NY: Routledge.

Gilbert, P. (2014). The origins and nature of compassion focused therapy. *British Journal of Clinical Psychology*, *53*(1), 6–41.

Gilbert, P., McEwan, K., Mitra, R., Franks, L., Richter, A., & Rockliff, H. (2008). Feeling safe and content: A specific affect regulation system? Relationship to depression, anxiety, stress, and self-criticism. *The Journal of Positive Psychology*, *3*(3), 182–191.

Gilbert, P., McEwan, K., Mitra, R., Richter, A., Franks, L., Mills, A., … Gale, C. (2009). An exploration of different types of positive affect in students and patients with bipolar disorder. *Clinical Neuropsychiatry*, *6*(4), 135–143.

Gillath, O., Hart, J., Noftle, E. E., & Stockdale, G. D. (2009). Development and validation of a state adult attachment measure (SAAM). *Journal of Research in Personality*, *43*(3), 362–373.

Halamová, J., Kanovskỳ, M., Gilbert, P., Troop, N. A., Zuroff, D. C., Hermanto, N., … Shahar, B. (2018). The factor structure of the Forms of Self-Criticising/Attacking & Self-Reassuring Scale in thirteen distinct populations. *Journal of Psychopathology and Behavioral Assessment*, *40*(4), 736–751.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Kabat-Zinn, J. (1982). An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: Theoretical considerations and preliminary results. *General Hospital Psychiatry*, *4*(1), 33–47.

Kelly, A. C., & Dupasquier, J. (2016). Social safeness mediates the relationship between recalled parental warmth and the capacity for self-compassion and receiving compassion. *Personality and Individual Differences*, *89*, 157–161.

Kelly, A. C., Zuroff, D. C., Leybman, M. J., & Gilbert, P. (2012). Social safeness, Received social support, and maladjustment: Testing a tripartite model of affect regulation. *Cognitive Therapy and Research*, *36*(6), 815–826. https://doi.org/10.1007/s10608-011-9432-5

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review, 110*(2), 265-284.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kocalevent, R.-D., Hinz, A., & Brähler, E. (2013). Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. *General Hospital Psychiatry, 35*(5), 551–555.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606–613.

Lavy, S., & Littman-Ovadia, H. (2011). All you need is love? Strengths mediate the negative associations between attachment orientations and life satisfaction. *Personality and Individual Differences*, *50*(7), 1050–1055.

Lau, M. A., Bishop, S. R., Segal, Z. V., Buis, T., Anderson, N. D., Carlson, L., … Devins, G. (2006). The Toronto Mindfulness Scale: Development and validation. *Journal of Clinical Psychology*, *62*(12), 1445–1467.

Liotti, G., & Gilbert, P. (2011). Mentalizing, motivation, and social mentalities: Theoretical considerations and implications for psychotherapy. *Psychology and Psychotherapy: Theory, Research and Practice*, *84*(1), 9–25.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*(3), 285–300.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care*, *46*(3), 266–274.

Mallinckrodt, B. (2010). The psychotherapy relationship as attachment: Evidence and implications. *Journal of Social and Personal Relationships, 27*(2), 262–270.

Mallinckrodt, B., & Jeong, and. (2014). Meta-Analysis of client attachment to therapist: Associations with working alliance and client pretherapy attachment. *Psychotherapy Theory Research & Practice*, *52*.

Martins, M. J., Marques, C., Carvalho, C. B., Macedo, A., Pereira, A. T., & Castilho, P. (2019). Engaging with the affiliative system through mindfulness: The impact of the different types of positive affect in psychosis. *Journal of Clinical Psychology*, *75*(3), 562–573.

McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Lawrence Earlbaum.

Pavot, W., & Diener, E. (2009). Review of the Satisfaction With Life Scale. In E. Diener (Ed.) *Assessing well-being* (pp. 101–117). New York, NY: Springer

R Core Team. (2017). *R: A Language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Revelle, W. (2018). psych: Procedures for personality and psychological research. Evanston, Illinois: Northwestern University https://CRAN.R-project.org/package=psych, Version = 1.8.12.

Rosseel, Y. (2012). Lavaan: An R *p*ackage for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36.

Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(39), 1161–1178.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78.

Shevlin, M., Brunsden, V., & Miles, J. N. V. (1998). Satisfaction With Life Scale: Analysis of

factorial invariance, mean structures and reliability. *Personality and Individual Differences*, *25*(5), 911–916.

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097.

Steiger, J., & Lind, J. (1980, May). *Statistically-based tests for the number of common factors.* Paper presented at the Annual Spring Meeting of the Psychometric Society, University of Iowa, Iowa City, IA.

Sterba, S. K., & MacCallum, R. C. (2010). Variability in parameter estimates and model fit across repeated allocations of items to parcels. *Multivariate Behavioral Research*, *45*(2), 322–358.

Teixeira, R. J., Ferreira, G., & Pereira, M. G. (2017). Portuguese validation of the Cognitive and Affective Mindfulness Scale-Revised and the Philadelphia Mindfulness Scale. *Mindfulness & Compassion*, *2*(1), 3-8.

Tejedor, R., Feliu-Soler, A., Pascual, J. C., Cebolla, A., Portella, M. J., Trujols, J., ... & Soler, J. (2014). Psychometric properties of the Spanish version of the Philadelphia Mindfulness Scale. *Revista de Psiquiatría y Salud Mental (English Edition)*, *7*(4), 157-165.

Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681-706). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form.*

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_h: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123–133.

Table 1

Demographic Information

	All	Completers
DEMOGRAPHIC INFO		
Education		
At least some university	93.5%	93.8%
Post-graduate or professional degree	54.3%	53.8%
Income		
>25k per year	80.4%	81.5%
>100k per year	39.1%	41.5%
STUDY VARIABLES	M (SD)	M (SD)
PA	1.96 (0.72)	1.90 (0.72)
NA	1.07 (0.66)	1.13 (0.72)
Social safeness	2.42 (0.81)	2.37 (0.77)
Anxiety symptoms	7.64 (4.42)	8.07 (4.50)
Depressive symptoms	6.47 (4.33)	6.89 (4.24)

Note: Demographic information for the full sample present at intake (N = 92) and reduced sample (N = 65) who went on to complete the exit measures.

Table 2

Change in Study Variables from Intake to Exit

VARIABLE	INTAKE	EXIT	DIFF	*t*	*p*
	M (*SD*)	*M* (*SD*)	*M* (*SD*)		
PA	1.90 (0.72)	2.20 (0.71)	.27(.61)	3.44	.001
NA	1.13 (0.72)	.71 (0.53)	-.40 (.65)	-4.86	<.001
Social safeness	2.37 (0.77)	2.71 (0.89)	.35 (.69)	3.73	<.001
Anxiety symptoms	8.07 (4.50)	4.59 (4.08)	-3.39 (4.22)	-6.38	<.001
Depressive symptoms	6.89 (4.24)	4.85 (4.32)	-1.91 (4.00)	-3.75	<.001

Note: Intake, exit, and difference means and standard deviations for all symptom and affect variables. Difference does not precisely match the difference of the reported intake and exit means as some scores were missing at each time point.

Table 3

Correlations among PA, NA, and Social Safeness at Intake.

	PA	NA	Safeness
PA	1	-.28*	.68***
NA	-.24*	1	-.48**
Safeness	0.61***	-0.45***	1

Note: Observed correlations are in the lower left. Correlations calculated using SEM are shaded and in the upper right.
*$p \leq .05$. **$p < .01$. ***$p < .001$.

Table 4

Bivariate and Unique Associations of Social Safeness with Personality Constructs at Intake

	Safeness	
	Bivariate	Unique
Self-criticism	-.66***	-.44**
Self-reassurance	.64***	.45*
Mindful - attention	.36**	.32
Mindful - acceptance	.20	.25
Life Satisfaction	.61***	.60**

Note. Bivariate and unique associations between latent social safeness and observed personality variables at intake. Unique associations are controlling for latent PA and NA. Fit for all models was adequate (CFI >.94, RMSEA <=.09). Factor loadings for all models were > .70.
*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$. +$.05 < p \leq .10$

Table 5

Bivariate and Unique Relationships of Social Safeness with Depression and Anxiety at Intake

	Depression	Anxiety
Safeness	-.52***	-.52***
PA	-.54***	-.46***
NA	.57***	.71***
Safeness (unique)	-.01	.02

Note. Bivariate associations between latent social safeness, PA and NA with observed depression and anxiety. Unique association for safeness is controlling for latent PA and NA. Models predicting anxiety symptoms (CFI = .95, RMSEA = 09) and depressive symptoms (CFI = .93, RMSEA = .11) both had adequate fit. Factor loadings for all models were > .70.
*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$. +$.05 < p \leq .10$

Table 6

Individual and Unique Contribution of Social Safeness in Predicting Anxiety and Depressive Symptoms

	Safeness	
	Individual	Unique
Depressive Symptoms	-.30*	-.01
Anxiety Symptoms	-.52***	-.72***

Note. Latent social safeness at the program midpoint predicting change in symptom outcomes: exit controlling for intake. 'Individual' includes no control variables. 'Unique' includes latent PA and NA, also at the program midpoint.
*$p \leq .001$.

General Discussion

Social safeness is the warm soothing experience of being connected to others. Theorized to have evolved alongside the attachment system, it is part of Gilbert's tripartite model of affect (Gilbert et al., 2008), which also includes PA and NA. Under this model, social safeness is believed to be extraordinarily important for emotion regulation and is taught explicitly in at least one form of therapy (Gilbert, 2009, 2014). The tripartite model stands in contrast to two-dimensional affective systems like valence and arousal (Russell, 1980), and Watson, Clark, and Tellegen's (1988) PA and NA model. The valence/arousal model posits that all self-reported affect exists in a two-dimensional space characterized by pleasantness-unpleasantness (valence) and high or low energy (arousal). Under this model, social safeness would merely be a location on this two-dimensional grid, and there would be little need to measure or assess it separately (Barrett & Russell, 1998; Posner, Russell, & Peterson, 2005; Russell, 1980). The creators of the PA and NA model make a similar but softer claim that activated emotion states have thus far been largely captured by two dimensions although additional dimensions may exist (Watson et al., 1988; Watson & Tellegen, 1985). The primary aim of this dissertation was to evaluate the case for including social safeness as a third dimension of affect alongside PA and NA.

In order to evaluate the independence of social safeness from PA and NA, four studies were conducted and organized into two articles. Studies One through Three are reported in Article One. Study Four is reported in Article Two. In this General Discussion, I will refer to the studies directly rather than referring to the article in which they appear. Study One was conducted cross-sectionally on Mechanical Turk using community participants. Studies Two and Three were conducted longitudinally in the lab and online with university students. Study Four was conducted as part of a psycho-educational intervention–Mindfulness Based Stress Reduction

(MBSR: Kabat-Zinn, 1982)–given at a local private psychology clinic. These four studies sought to evaluate the independence of social safeness in three primary ways: (1) by factor analytically distinguishing PA, NA and safeness items, (2) by predicting social safeness with PA and NA and estimating the percentage of variance in safeness they explained, and (3) by examining the unique predictive power of social safeness in predicting other constructs above and beyond PA and NA.

Factor Analysis

In order to establish that the PA, NA, and safeness items were factorially distinct, an exploratory factor analysis (EFA) of the combined set of items was conducted in one half of the data in Study One. The resulting three-factor structure was then tested using confirmatory factor analysis (CFA) in the second half of the Study One data and in all subsequent studies. This was done in order to establish that the item-level variance of the measures was best explained by three distinct factors representing the three constructs in question and to validate the measurement model for further analyses. If social safeness did not separate into a clear third factor, it would have been difficult to argue that it was distinct from the other two affective dimensions. In particular, the fact that social safeness shares a pleasant, positive valence with PA made it plausible that their items would have been best explained by a single factor or would not have separated cleanly (i.e., with no cross-loadings) into two factors. However, PA, NA and social safeness separated cleanly into three distinct factors with no substantial cross-loadings in EFA, and in CFA the three-factor model fit well in all four studies with one minor exception: the addition of a covariance was required between two of the safeness items in Study Four for the program midpoint data, but not the intake data. The results of the EFA in Study One and the CFAs in all four studies provided strong evidence that the social safeness items in the TPAS are

factorially distinct from the PA and NA items in the PANAS. This was particularly important given that one of our aims was to distinguish these constructs using instruments that were as similar in format as possible.

Predicting Social Safeness with PA and NA

After testing the factorial distinctness of social safeness, we sought to examine the relationships among PA, NA and social safeness and to address the argument that one's standing on social safeness could be fully predicted by measuring PA and NA, as this would mean that social safeness was redundant with those constructs. In order to investigate this, we first examined the latent bivariate associations among the three constructs in SEM. We then regressed social safeness on PA and NA in order to examine their unique relationships with safeness and the total percentage of variance they explained in safeness. This was done in all four studies. In Studies Two and Three, which were longitudinal, the models were multi-level, allowing us to examine the constructs both between-persons and within-person.

PA and NA had substantial bivariate and unique relationships with social safeness cross-sectionally in Studies One and Four, and between- and within-person in Studies Two and Three. The relationships with PA were positive and medium to large in size. The relationships with NA were negative and medium to large in size. These relationships were found at both the between- and within-person levels and remained as unique relationships when PA and NA were both included in the regression. These associations were in line with prior work examining the relationships among these constructs (Kelly & Dupasquier, 2016; Kelly, Zuroff, Leybman, & Gilbert, 2012) and with the theoretical assertion that social safeness soothes and deactivates feelings of distress/NA and supports feelings of engagement/PA (Gilbert, 2014).

The percentage of variance explained in social safeness by PA and NA varied from study to study, ranging from as low as 35% of the within-person variance in Study Three to as high as 61% of the between-persons variance in Study Two. The value was always well below the statistical threshold for variable redundancy of 80-90% (Kleinbaum, Kupper, Nizam, & Rosenberg, 2013), and in many cases it was lower than even a simple majority. As these models were all conducted using latent variables in SEM, the variance not accounted for in social safeness could not be attributed to random error but was instead systematic and unique to safeness. It was concluded that while PA and NA accounted for a substantial percentage of the variance in social safeness, a person's standing on safeness could not be determined solely by measuring PA and NA. Combined with the factorial distinction between the constructs, this was taken as strong support for safeness's independence.

Predicting Outcomes with Social Safeness Controlling for PA and NA

After establishing the factorial and analytic independence of social safeness, we sought to investigate its utility in predicting other constructs. For social safeness to be useful to clinicians and researchers it needs not only to be independent of PA and NA but also to be predictive of other relevant constructs. If safeness could be shown to be uniquely predictive even when controlling for PA and NA, this would be evidence both for independence and utility. Throughout the four studies, constructs with previously identified theoretical relationships with social safeness were chosen as part of an effort to validate the larger theoretical model set forth in CFT. This model includes not only PA, NA and safeness but also self-compassion/self-reassurance, self-criticism, and support processes related to attachment (Gilbert, 2014). These analyses began in Study Two as Study One did not include any additional constructs.

Social safeness was found to have a medium-sized, unique negative association with perceived stress at both the between- and within-person levels in Study Two. This was expected as social safeness is conceptualized as a system that soothes distress (Gilbert et al., 2008). Social safeness also had substantial unique positive relationships with several support related constructs in Study Three: perceived, received and given support. As self-reassurance can be thought of as self-directed caregiving (Hermanto & Zuroff, 2016), we include it here as a fourth support construct, *self-support*. With the exception of perceived support, these relationships were found only at the within-person (daily) level. These within-person support relationships extend the work of Kelly et al. (2012) by including given support and self-support (self-reassurance) and by showing the unique contribution of social safeness above and beyond PA and NA. Our study also used the adjective-based measure of social safeness which I believe serves as a better state-level measure of the construct. These relationships with social support supported the hypothesized relationship between social safeness and support processes related to attachment. The relationship with self-reassurance in particular supports the CFT model's assertion that one can activate one's own social safeness system through self-reassurance/self-support (Gilbert, 2009, 2014). Interestingly, social safeness had a very large between-person association with perceived support, so high that one might argue that the two constructs represent the same underlying phenomenon. This finding supported safeness's strong theoretical relationship with attachment security and perceived support availability. That this relationship was much smaller within-person (r = .21) highlights how social safeness differs from those constructs at the daily level. It is as the time scale moves closer to the moment that the felt affective quality of social safeness becomes more apparent.

Study Four sought to extend previous work by examining for the first time the relationship of social safeness with change in clinical symptoms over the course of a psychological intervention. This was only the second time that social safeness had been measured in a population with elevated clinical symptoms. The relationship of social safeness with self-criticism and self-reassurance, which had been shown in previous work (Gilbert et al., 2008; Kelly et al., 2012; Richter, Gilbert, & McEwan, 2009), was here shown to hold controlling for PA and NA. Study Four was also the first time that the relationship of social safeness was examined with life satisfaction or the attention control aspect of mindfulness. While there was no unique relationship with either component of mindfulness, social safeness did show a strong unique relationship with life satisfaction. Social safeness may be an important indicator of the more general construct subjective well-being, which is a composite of life satisfaction, PA, and NA (Diener, 2009).

Study Four replicated the relationships of social safeness with anxiety and depression shown in previous work (Alavi et al., 2016; Gilbert et al., 2008; Richter et al., 2009). Interestingly, neither of these relationships held controlling for PA and NA. Given the substantial overlap of PA with depression and of NA with anxiety (Clark & Watson, 1991), it is possible that there was simply too little variance left for safeness to remain predictive. No previous studies have examined the relationship between social safeness and anxiety symptoms controlling for PA and NA, and only one previous study has examined safeness's relationship with depressive symptoms with these controls (Kelly et al., 2012). Contrary to our findings, in this earlier work the unique relationship with depressive symptoms had remained significant. One reason for this difference might be the use of different depressive symptom measures. Kelly et al. employed the Beck Depression Inventory (Beck, Steer & Brown, 1996), which includes elements related to self-criticism (e.g., "I believe that I look ugly" and "I hate myself.") that are

not present in the Patient Health Questionnaire (Kroenke, Spitzer, & Williams, 2001), which was used in Study Four. Social safeness may have uniquely predicted depressive symptoms in that sample as a result of its relationship with self-criticism (Gilbert et al., 2008; Kelly et al., 2012).

Social safeness also significantly predicted *change* in depressive and anxiety symptoms and while the relationship with change in depressive symptoms did not hold controlling for PA and NA, the effect for anxiety did. This effect for change in anxiety symptoms may have held, because social safeness is fundamentally a soothing system rather than an energizing one (Gilbert, 2014). Anxiety is characterized by activated distress, while depression is characterized by a lack of activation and enthusiasm (Clark & Watson, 1991), which would leave little to soothe/deactivate. If reductions in depressive symptoms over the course of the intervention were driven by increases in motivation, the PA component of depression, little change variance may have remained when PA was controlled. Another possible explanation for the presence of an effect for anxiety change but not depressive symptom change is the difference in the size of the two effects. The anxiety change effect was roughly twice the size of the change in depressive symptoms, both in raw and standard deviation units. The reduced range for depressive symptom change may account for the reduced social safeness effect.

Theoretical Implications

The work presented in this dissertation has contributed to psychological research in several areas. It has shown in detail the psychometric properties of social safeness as measured by the TPAS. More importantly, it has clarified the relationship of the TPAS safeness adjectives to the similarly constructed PA and NA adjectives. This has important implications for the measurement of social safeness and potentially for the structure of self-reported affect more generally. The present work has also drawn important links between social safeness and support

and care processes that are important for illuminating the relationship between social safeness and attachment. Finally, social safeness was measured alongside self-reassurance and self-criticism, which provided empirical validation for important parts of the CFT model. This, combined with the effects found for anxiety and depressive symptoms and changes in those symptoms during an intervention, provided strong support for the clinical relevance of social safeness.

Social safeness and the structure of affect. Gilbert initially conceived of social safeness as a third affective dimension alongside PA and NA (Gilbert et al., 2008), and we have presented compelling evidence that it is indeed independent. By showing that at least three basic dimensions are needed to properly characterize affect, we are showing not just that, "It's three not two," but that a more complicated structure of affect may be necessary. Tellegen et al., (1999) suggested that there are three hierarchical levels in the structure of self-reported affect, with a bipolar valence dimension at level three, PA and NA at level two, and individual affects at level one. The present work, and Gilbert's tripartite model more generally, suggest that a social safeness factor is necessary at level two of that hierarchy. See Figure 1 for a visual depiction of social safeness's potential location in that model. I have chosen to represent this location with an empty box, because I believe there are several possibilities for how social safeness may be structured. I lay out three possibilities in Figure 2. The first is a single emotion-behavior system residing at level two. This would stand in contrast to PA and NA which are higher order dimensions that subsume individual emotion systems. Placing social safeness in this position would be somewhat inconsistent with the other dimensions at this level, but it is possible that nature is factorially inconsistent in this way.

The second possible structure places social safeness at level two and has it comprised of individual emotions systems. This would put social safeness at the same hierarchical level as PA and NA, and it would also change the nature of the social safeness construct somewhat. Currently safeness is conceived as a single felt affect rather than a family of affects, so this would represent a change from existing theory (Gilbert, 2014; Gilbert et al., 2008). The individual emotions that I propose for structure two include, but are not necessarily limited to: experienced emotional warmth, contentedness, and a feeling of connection with others. These three individual emotions were chosen based on the social safeness item content and my own intuition. A more rigorous examination of social safeness and its possible components would be needed to establish this possible structure properly.

The third possible structure places social safeness at level one as an individual emotion-behavior system. This position would be more consistent with the PA/NA model which places individual emotions at level one. However, this structure would require the creation of a new overarching dimension of which social safeness would be a component. I have chosen to call this dimension *affiliation*. I propose that it includes, but is not necessarily limited to: grief/loss, social safeness, and compassion. Grief/loss was chosen both because it represents the attachment related emotion system associated with care-seeking in the Panksepp (2010) model and because it has been suggested that grief/loss represents a process akin to withdrawal from an addictive substance, in this case the neurochemicals associated with love and attachment (Panksepp, 1998, 2010). As such it is plausible that grief/loss and social safeness exist as part of a higher order emotional framework representing care-seeking and the capacity to be soothed by care receipt respectively. I included compassion as the third individual emotion in this set as Gilbert has suggested that compassion represents a specific motivation-behavior system (Gilbert, 2014), the

motivated desire to help, that Gilbert views as related to social safeness (Gilbert, 2014). Compassion is associated with much of the same neural circuitry as safeness (e.g., Rockliff et al., 2011), but it specifically motivates care-giving rather than care-seeking or the capacity to be soothed. Thus, it would make sense that compassion would have a felt affective experience distinct from those associated with those other care processes. To summarize, the higher order affiliation dimension would be comprised of: (1) the distress associated with separation and need for care, which motivates care-seeking, (2) the warm felt security that soothes and connects when care or proximity are experienced, and (3) the compassionate feelings that motivate caregiving.

While the proponents of the PA/NA model are open to the possibility of additional dimensions (Watson et al., 1988; Watson & Tellegen, 1985), the proponents of the valence/arousal model are quite committed to two basic dimensions (Posner et al., 2005). Since its inception, the valence/arousal model has developed into Barret's *theory of constructed emotions* (2006). Under this model, valence and arousal are not merely ways of characterizing self-reported affect; instead they represent the basic components of which affects are constructed. All other distinctions (e.g., fear and anger, determination and joy) are considered to be cognitive elaborations that occur downstream of the more basic valence/arousal neural circuitry (Barrett, 2017). This claim rests, in part, on the theory that human emotions exist on a two-dimensional circumplex. That is, all basic emotions are arranged in a circle based on their levels of valence and arousal. See Figure 3. I would argue that by supporting the tripartite model of affect, we are providing evidence *against* a two-dimensional affective circumplex, though not necessarily against constructed emotions. An independent social safeness means that PA, NA and safeness cannot be well-represented in two-dimensional space. In particular, social safeness and low NA, which we have shown to be distinct, occupy the same position in the

valence/arousal grid, the lower right-hand corner, and cannot therefore be meaningfully distinguished by their standing on valence and/or arousal. This is not to say that valence and arousal are not represented in the brain or even that they are not basic components of emotional experience as there is much research supporting these claims (Barrett, 2006, 2017). Instead our evidence merely suggests that valence and arousal are not sufficient in themselves to fully characterize the wide variety of self-reported affect.

Social safeness, social support and attachment theory. Social safeness was conceived as an affective system that evolved in tandem with the attachment system and ultimately grew to include experiences and relationships outside of attachment (Gilbert, 2014; Gilbert et al., 2008). As such, it has theoretical and empirical associations with caregiving and care-seeking (Gilbert, 2014; Gilbert et al., 2008; Kelly et al., 2012). An important contribution of the present work was to distinguish between the felt experience of social safeness in-the-moment (within-person), where it was related to the actual transaction of support (given and received support), and the general tendency to experience social safeness (between-person), where it had a very strong relationship with perceived support availability. Both of these findings are what one would expect if social safeness is the emotion system underlying attachment security. In-the-moment it is the warm soothing feeling of connecting with others and in general it is the tendency to experience this warm feeling and the sense that it is available if needed.

As mentioned in the General Introduction, while PA and NA each have personality correlates at the trait level (extraversion and neuroticism, respectively), no such association has been suggested for social safeness. Attachment security is a likely candidate, and while we did not measure attachment security directly, we did find a very high correlation with perceived social support, a construct that is so highly related to attachment security that it has been argued

that they are in fact the same construct (Moreira et al., 2003). Taken together with the large associations found by Kelly et al. (2012) between social safeness and both perceived support *and* attachment security we now have growing support for attachment security/perceived support as the personality correlate of social safeness.

If attachment security (or perceived support) is the personality trait correlate of social safeness, safeness may be the state-level affective component of attachment security. Currently, although attachment security does have a state-level measure (Gillath, Hart, Noftle, & Stockdale, 2009), the measure does not focus on affective experience. Thus, there is currently no state-level measure of the affective experience of attachment security. Given the strong theoretical and empirical relationships between social safeness and attachment security, I would suggest that the state measure of affective attachment security *is* social safeness and that attachment researchers might therefore consider using social safeness as a self-report measure of this construct.

Self-reassurance, self-criticism, and the CFT model. In CFT, helping clients activate social safeness is one of the primary aims of therapy (Gilbert, 2009, 2014). Safeness is viewed as a fundamental emotion regulation tool and one that can be self-activated through the use of self-reassurance/self-compassion techniques. In addition, self-reassurance/self-compassion is viewed as a kind of antidote to self-criticism, and this process is thought to be mediated by social safeness (Gilbert, 2014). Thus, the CFT model predicts that social safeness has strong relationships, in opposite directions, with self-reassurance and self-criticism. It is also important that these relationships hold controlling for PA and NA. If, for example, the relationship between social safeness and self-criticism were fully explained by a reduction in NA, it would call into question the important theoretical role of safeness and perhaps even self-compassion in CFT. Thus, by showing that social safeness had a unique relationship with both self-reassurance and

self-criticism and that this relationship held controlling for PA and NA, we provided important additional support for the CFT model.

Future Directions

The introduction of a third fundamental affective dimension raises important questions about the structure of human affective experience. That the exact nature of this dimension is unclear is also an important gap worthy of attention. Regardless of where safeness resides in the structure of affect, it potentially represents a state measure of affective attachment security which should bear on research in attachment. Finally, we have shown that social safeness has an important relationship with clinical symptom outcomes. Combined with support for the CFT model, this is evidence that safeness's role in a clinical setting deserves additional attention.

Structure of affect. The current work established that social safeness does not fit well into existing two-dimensional affect models. In particular, our results do not appear to be compatible with the affective circumplex, and they indicate the need for an additional level two construct in Watson et al.'s affect hierarchy (1999). Two important research questions arise from these observations: the first related to the affective circumplex and the second to the affect hierarchy. We have suggested that social safeness is incompatible with the valence and arousal circumplex, but we did not test this directly. In order to solidify this claim, one would need to measure social safeness alongside valence and arousal and perform similar analyses to those in the present work. While we would not expect the results of such analyses to differ from our results with NA and PA, this is an empirical question that must be tested. This could be part of a larger work examining the ability of the valence and arousal grid to predict standing on a variety of self-reported affective dimensions including but not limited to social safeness. *Dominance* has been suggested as such a dimension (Mehrabian, 1980), and I might suggest *affiliativeness* as

another. Thus, if we added both potential dimensions, an emotional state could potentially be positive or negative, high or low energy, dominant or submissive, and socially oriented (affiliative) or not. The predictive validity of any additional fundamental dimensions of affect at this same level could also be investigated in such a study.

The second question of affect structure is related to the three proposed structures of social safeness mentioned above. While we showed that an additional structure is necessary at level two of the affect hierarchy, the nature of this structure is unclear. Social safeness may be an individual behavior-emotion system, a higher order affective dimension, or a combination of the two (i.e., an individual emotion that resides at level two rather than level one). Testing whether social safeness is composed of lower-order individual emotions would likely require a new measurement instrument. It would not be possible to distinguish subcomponents in the TPAS as it has only four items, and the content of the SSPS is largely cognitive rather than affective in nature. Investigating whether social safeness exists alongside other affects (e.g., grief/loss and compassion as I suggested) as part of an as yet unidentified higher order dimension would require a variety of different types of evidence. Additional candidate emotions for this affective family would need to be established, and self-report measures of a similar item type (adjectives) would need to be created as currently no such measures exist for grief or compassion. The same would also be true for any additional affect candidates. As the individual emotions subsumed by the current affective dimensions, PA and NA, share neural circuitry and neurotransmitter systems, a similar neurobiological link could be sought between safeness and any closely related affects that were identified.

To further investigate social safeness's place in the affective hierarchy and its precise relationships with other constructs, additional evidence could be collected for social safeness at

the momentary level. To date, all investigations of state-level social safeness have been conducted at the daily level, but it is possible that social safeness would have a different relationship with PA and/or NA when measured in-the-moment. To precisely understand the relationship of social safeness with other affects and the overall affective hierarchy in general, ecological momentary assessment (EMA) will be required. As these techniques typically use very brief measures of the target construct, appropriate measures of PA, NA, safeness and any other constructs to be tested (e.g. grief or compassion) will need to be gleaned from previous work or created.

Attachment. Social safeness has a close theoretical and empirical relationship with attachment security. However, this relationship has only been empirically measured twice (Gilbert et al., 2008; Kelly et al., 2012) and never at the state or momentary level. Thus, an important next step in this area might be to investigate the relationship between social safeness and attachment dimensions (anxiety, avoidance, and security) using a variety of instruments and time frames. If this work were conducted using a longitudinal design allowing measurements both between- and within-person, this would allow testing of the assertion that social safeness is strongly related to attachment security at the trait level and felt security at the state-level. Ideally research of this type would be conducted using both daily measures of the target constructs as well as EMA techniques. Daily measures can be easily adapted from existing instruments making results more consistent with existing research, and EMA allows researchers investigate in-the-moment relationships without resorting to more error-prone retrospective instruments. As another conceptual difference between social safeness and these attachment constructs is whether their scope includes non-attachment figures, alternative versions of daily or EMA instruments could be employed that widen or narrow the scope. For example, social safeness items could be

given with the instruction, “when I’m with my romantic partner” or attachment security items could be given with the instruction, “in any of my relationships.” By varying the instructional context in this way, one could empirically test the extent to which social safeness fits into an attachment framework. This would be an important first step toward the potential use of the social safeness construct in attachment research and greater communication between affect and attachment researchers.

Clinical importance. Social safeness is an integral part of the CFT model, and the relationships with self-related treatment constructs (e.g., self-criticism) and clinical outcomes (e.g., anxiety symptoms) reported here and elsewhere support its importance for emotion regulation in and outside of a clinical setting. The finding in Study Four that social safeness is predictive of clinical symptom change, which had previously been only a theoretical assertion (Gilbert, 2014), support it as a potential mechanism for change in psychotherapy. That said, the evidence from Study Four was preliminary. We measured change in the context of a particular intervention, MBSR groups, and with a particular population, predominantly female middle-class individuals with elevations in anxiety and depressive symptoms. Similar studies will need to be conducted in the context of other diagnostic groups, demographic populations, and intervention types in order to establish whether social safeness plays a similar role in other contexts or whether our findings were specific to our sample. In discussing why our depression finding did not hold controlling for PA and NA while the finding of Kelly et al. (2012) did, I suggested that this was due to the difference in depression measures and their differing symptom profiles. Future work will need to employ a variety of symptom measures that come from different theories of psychopathology in order to verify the scope of social safeness’s influence on clinical outcomes. Finally, while social safeness has associations with a variety of indicators of

psychopathology beyond depressive and anxiety symptoms (e.g., Gilbert et al., 2009; Kelly et al., 2012), the relationship of social safeness to *change* in these indicators will need to be investigated if we are to understand the broader relationship between social safeness and clinical change.

The mechanisms by which social safeness leads to clinical change are also worth investigating. In the CFT model, therapy focuses primarily on cultivating compassion for the self in order to generate feelings of warmth and safeness, which are seen as therapeutic. The extent to which feelings of safety experienced during therapy sessions are instrumental in this process remains to be determined. Therapy may merely provide a convenient space to practice self-compassion (or other) techniques, and safeness experienced in therapy may only function as an indicator that a technique is working. Another possibility is that safeness experienced in the therapy room may support and encourage the client to engage with the other techniques and exercises associated with therapy (e.g., mindfulness techniques in MBSR), but it is these other techniques that are responsible for symptom change. Finally, the experience of social safeness with the therapist may be an important part of raising the client's general level of safeness first in and then outside of therapy, and it is this general level of safeness that drives symptom change.

According to the emotion-focused therapy model, by providing the Rogerian conditions of positive regard, empathic attunement, and congruence (Rogers, 1957), the therapist is able to soothe and connect with the client in therapy until ultimately this "interpersonal regulation of affect becomes internalized into self-soothing and the capacity to regulate inner states" (Greenberg, 2014, p. 351). This interpersonal regulation of self-soothing affect sounds remarkably like social safeness. I would therefore hypothesize that social safeness is a key mechanism by which a warm therapeutic relationship based on the Rogerian conditions produces

outcomes. This could be assessed empirically by taking measures of the therapeutic relationship (e.g., Gelso et al., 2005), social safeness, and clinical outcomes over the course of a therapy intervention. Simply examining whether safeness assessed in-session is predictive of measures of the therapeutic relationship and/or treatment outcomes would be an important first step to understanding safeness's therapeutic role. If social safeness is a mechanism of relationship-based change, one would expect a more general measure of social safeness experienced over the course of therapy to at least partially mediate the association between the therapeutic relationship and treatment outcomes. A more complicated mediation model that also includes general levels of social safeness outside of therapy could also be investigated.

Conclusion

Social safeness is the warm soothing affect that is thought to drive affiliation and attachment processes. As part of Gilbert's tripartite model of affect, it stands beside PA and NA as one of three core affective processes relevant in and outside of therapy. In the present work, I investigated whether social safeness is an independent third dimension of affect, distinct from PA and NA. I showed that social safeness is factorially distinct. Its variance cannot be fully explained by PA and NA, and it uniquely predicts personality constructs and clinical symptom change in anxiety even when controlling for PA and NA. This work supports the CFT model by showing that social safeness has the theoretically predicted relationships with self-reassurance, self-criticism, and social support. Taken together, I have presented strong evidence that social safeness is indeed an affect in its own right, distinct from PA and NA and that it is highly relevant in the areas of personality, attachment, affect, and clinical outcome research. I would strongly suggest that any researcher in these areas consider adding the four social safeness adjective items to any future assessment package they employ.

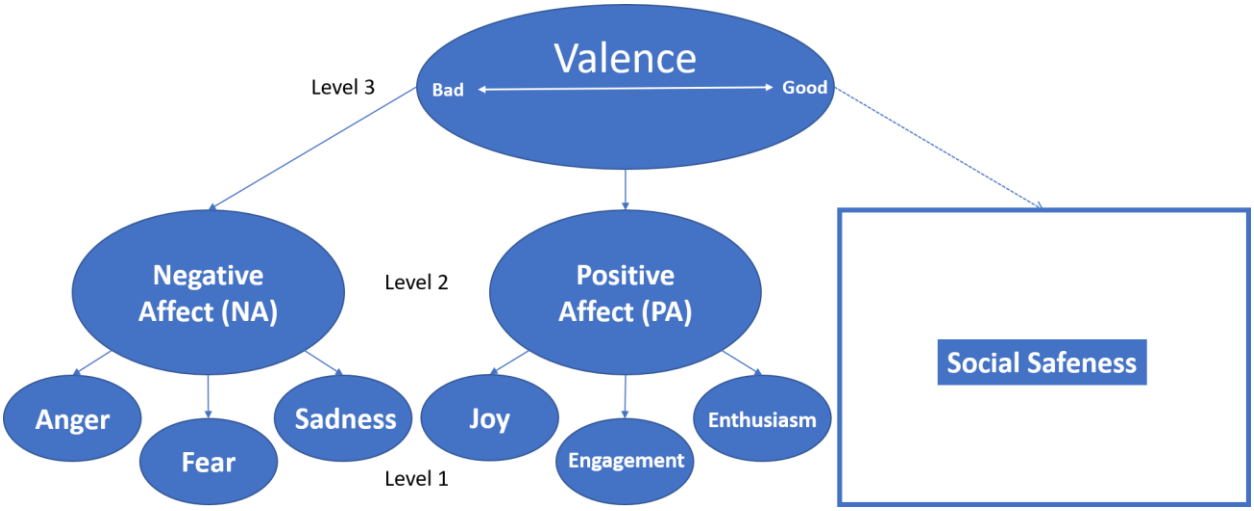

Figure 1. Structure of affect including social safeness.
The three-level structure of affect according to Watson et al., (1999) is depicted to the left with the proposed addition of social safeness to the right. Only three individual emotions out of many are shown for PA and NA for simplicity. Social safeness is left blank as there are multiple possible structures (See Figure 2).

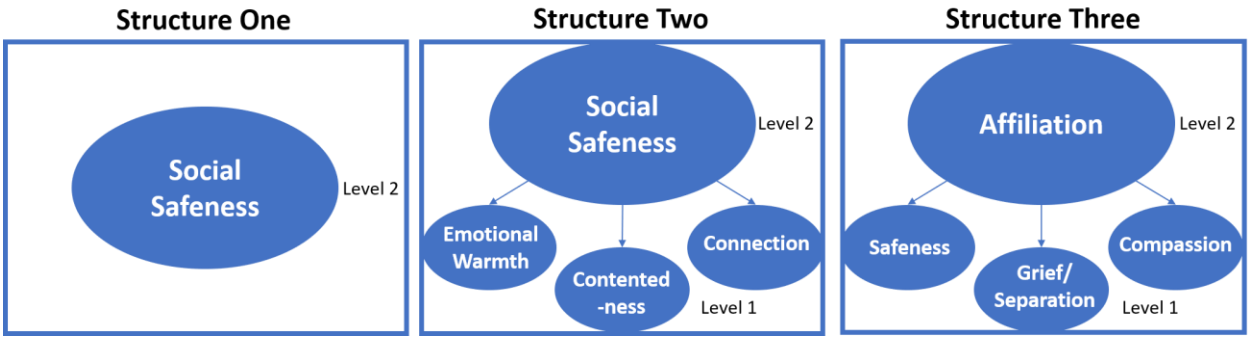

Figure 2. Alternate structures of social safeness.
Three potential structures for social safeness are depicted. Structure One represents safeness as a single emotion system occupying a level two position but subsumes no lower-order emotions. Structure Two represents social safeness as an affective dimension at level two comprised of several level one emotion systems. Structure Three represents social safeness as a level one emotion system that is part of a more general level two affective dimension called "affiliation". Only three individual (level one) emotions are included for Structure Two and Structure Three for simplicity. The existence of additional and/or different individual emotions is possible.

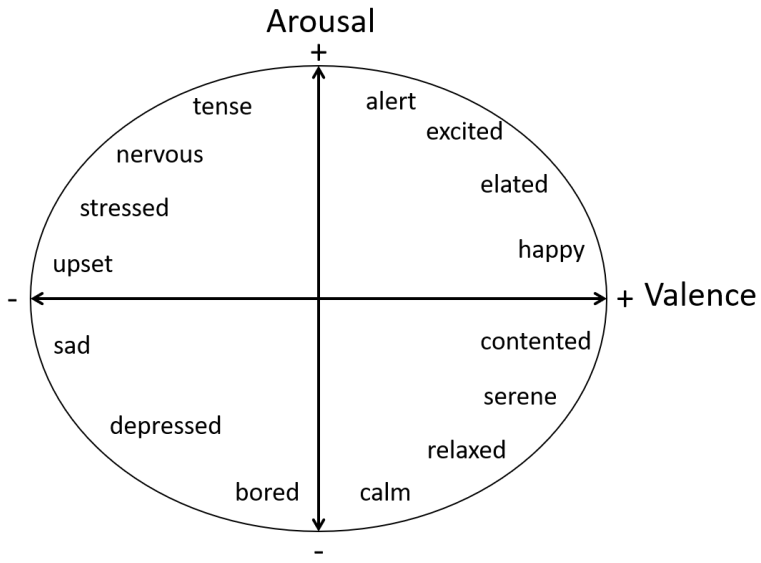

Figure 3. The affective circumplex.
Here is represented the valence (pleasant-unpleasant) and arousal (activation-deactivation) circumplex (adapted from: Russell, 1980) whereby individual emotions exist in a circle based on their relative standings on the two dimensions.

General References

Ainsworth, M. D., & Bell, S. M. (1970). Attachment, exploration, and separation: Illustrated by the behavior of one-year-olds in a strange situation. *Child Development*, *41*(1), 49–67.

Akin, A., & Akin, U. (2015). Self-compassion as a predictor of social safeness in Turkish university students. *Revista Latinoamericana de Psicología*, *47*(1), 43–49.

Alavi, K., Moghadam, A., Ali, M., Rahiminezhad, A., Farahani, H., & Modares Gharavi, M. (2016). Psychometric properties of Social Safeness and Pleasure Scale (SSPS). *Journal of Fundamentals of Mental Health*, *19*(1), 5–13.

Barrera, M. (1986). Distinctions between social support concepts, measures, and models. *American Journal of Community Psychology*, *14*(4), 413–445.

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, *10*(1), 20–46.

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23.

Barrett, L.F., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, *74*(4), 967.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529.

Beck, A. T., Steer, R. A., Brown, G. K. (1996) *BDI-II, Beck depression inventory: manual.* San Antonio, Texas: Psychological Corp.

Blankstein, K. R., Dunkley, D. M., & Wilson, J. (2008). Evaluative concerns and personal standards perfectionism: Self-esteem as a mediator and moderator of relations with personal and academic needs and estimated GPA. *Current Psychology*, *27*(1), 29–61.

Blatt, S. J. (2007). A fundamental polarity in psychoanalysis: Implications for personality development, psychopathology, and the therapeutic process. *Psychoanalytic Inquiry*, *26*(4), 494–520.

Blatt, S. J., D'Afflitti, J. P., & Quinlan, D. M. (1976). Experiences of depression in normal young adults. *Journal of Abnormal Psychology*, *85*(4), 383.

Blatt, S. J., & Zuroff, D. C. (1992). Interpersonal relatedness and self-definition: Two prototypes for depression. *Clinical Psychology Review*, *12*(5), 527–562.

Bourdieu, P. (1986). *The forms of capital*. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education*: 241-258. New York: Greenwood.

Bowlby, J. (1969). *Attachment and loss v. 3 (Vol. 1)*. Random House

Furman, W., & Buhrmester, D.(2009). Methods and Measures: The Network of Relationships Inventory: Behavioral Systems Version. *International Journal of Behavioral Development*, *33*, 470–478.

Bowlby, John. (1980). *Attachment and loss*. New York, NY US: Basic Books.

Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement of adult attachment: An integrative overview. In *Attachment theory and close relationships* (pp. 46–76). New York, NY, US: Guilford Press.

Carver, C. S. (1997). Adult attachment and personality: Converging evidence and a new measure. *Personality and Social Psychology Bulletin*, *23*(8), 865–883.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100*(3), 316.

Cohen, S. (2004). Social relationships and health. *American Psychologist*, *59*(8), 676.

Cohen, S., & Hoberman, H. M. (1983). Positive events and social supports as buffers of life change stress. *Journal of Applied Social Psychology*, *13*(2), 99–125.

Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin*, *98*(2), 310–357.

Collins, N. L., & Read, S. J. (1990). Adult attachment, working models, and relationship quality in dating couples. *Journal of Personality and Social Psychology*, *58*(4), 644.

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, *114*(38), E7900–E7909.

Crockett, K. B., & Turan, B. (2018). Moment-to-moment changes in perceived social support and pain for men living with HIV: An experience sampling study. *PAIN*, *159*(12), 2503–2511.

Cutrona, C. E., & Russell, D. W. (1987). The provisions of social relationships and adaptation to stress. *Advances in Personal Relationships*, *1*(1), 37–67.

Davis, K. L., Panksepp, J., & Normansell, L. (2003). The affective neuroscience personality scales: Normative data and implications. *Neuropsychoanalysis*, *5*(1), 57–69.

Depue, R. A., Luciana, M., Arbisi, P., Collins, P., & Leon, A. (19950201). Dopamine and the structure of personality: Relation of agonist-induced dopamine activity to positive emotionality. *Journal of Personality and Social Psychology*, *67*(3), 485.

Depue, R. A., & Morrone-Strupinsky, J. V. (2005). A neurobehavioral model of affiliative bonding: Implications for conceptualizing a human trait of affiliation. *Behavioral and Brain Sciences*, *28*(3), 313–350.

Diener, E. (2009). Subjective well-being. In *The science of well-being* (pp. 11–58).

Dunkel-Schetter, C., & Bennett, T. L. (1990). Differentiating the cognitive and behavioral aspects of social support. In B. R. Sarason, I. G. Sarason, & G. R. Pierce (Eds.), *Social support: An interactional view.* (pp. 267–296). Oxford England: John Wiley & Sons.

Dunkley, D. M., Zuroff, D. C., & Blankstein, K. R. (2003). Self-critical perfectionism and daily affect: Dispositional and situational influences on stress and coping. *Journal of Personality and Social Psychology*, *84*(1), 234.

Ekman, P., & Keltner, D. (1997). Universal facial expressions of emotion. *Segerstrale U, P. Molnar P, Eds. Nonverbal Communication: Where Nature Meets Culture*, 27–46.

Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, *74*(4), 967.

Fraley, R. C., Waller, N. G., & Brennan, K. A. (20000301). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology*, *78*(2), 350.

Gelso, C. J., Kelley, F. A., Fuertes, J. N., Marmarosh, C., Holmes, S. E., Costa, C., & Hancock, G. R. (2005). Measuring the real relationship in psychotherapy: Initial validation of the therapist form. *Journal of Counseling Psychology*, *52*(4), 640.

Gilbert, P., Cheung, M. S.-P., Grandfield, T., Campey, F., & Irons, C. (2003). Recall of threat and submissiveness in childhood: Development of a new scale and its relationship with

depression, social comparison and shame. *Clinical Psychology & Psychotherapy*, *10*(2), 108–115.

Gilbert, P. (2005). Compassion and cruelty: A biopsychosocial approach. In *Compassion* (pp. 21–86). Routledge.

Gilbert, P. (2009a). Introducing compassion-focused therapy. *Advances in Psychiatric Treatment*, *15*(3), 199–208.

Gilbert, P. (2009b). *The compassionate mind*. Robinson.

Gilbert, P. (2014). The origins and nature of compassion focused therapy. *British Journal of Clinical Psychology*, *53*(1), 6–41.

Gilbert, P. (2015). An evolutionary approach to emotion in mental health with a focus on affiliative emotions. *Emotion Review*, *7*(3), 230–237.

Gilbert, P., Clarke, M., Hempel, S., Miles, J. n. v., & Irons, C. (2004). Criticizing and reassuring oneself: An exploration of forms, styles and reasons in female students. *British Journal of Clinical Psychology*, *43*(1), 31–50.

Gilbert, P., McEwan, K., Mitra, R., Franks, L., Richter, A., & Rockliff, H. (2008). Feeling safe and content: A specific affect regulation system? Relationship to depression, anxiety, stress, and self-criticism. *The Journal of Positive Psychology*, *3*(3), 182–191.

Gilbert, P., McEwan, K., Mitra, R., Richter, A., Franks, L., Mills, A., … Gale, C. (2009). An exploration of different types of positive affect in students and patients with bipolar disorder. *Clinical Neuropsychiatry*, *6*(4), 135–143.

Gillath, O., Hart, J., Noftle, E. E., & Stockdale, G. D. (2009). Development and validation of a state adult attachment measure (SAAM). *Journal of Research in Personality*, *43*(3), 362–373.

Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology*, *64*(6), 1029–1041.

Greenberg, L. (20140728). The therapeutic relationship in emotion-focused therapy. *Psychotherapy*, *51*(3), 350

Griffin, D. W., & Bartholomew, K. (1994). Models of the self and other: Fundamental dimensions underlying measures of adult attachment. *Journal of Personality and Social Psychology*, *67*(3), 430.

Hammen, C. (2005). Stress and Depression. *Annual Review of Clinical Psychology*, *1*(1), 293–319.

Hazan, C., & Shaver, P. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology*, *52*(3), 511–524.

Hermanto, N., & Zuroff, D. C. (2016). The social mentality theory of self-compassion and self-reassurance: The interactive effect of care-seeking and caregiving. *The Journal of Social Psychology*, *0*(0), 1–13.

Horvath, A. O., & Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy. *Journal of consulting and clinical psychology*, *61*(4), 561.

Jacobs, J. (1961). The death and life of great american cities. *Vintage, New York.*

Kabat-Zinn, J. (1982). An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: Theoretical considerations and preliminary results. *General Hospital Psychiatry*, *4*(1), 33–47.

Kelly, A. C., & Dupasquier, J. (2016). Social safeness mediates the relationship between recalled parental warmth and the capacity for self-compassion and receiving compassion. *Personality and Individual Differences*, *89*, 157–161.

Kelly, A. C., Zuroff, D. C., Leybman, M. J., & Gilbert, P. (2012). Social safeness, received social support, and maladjustment: Testing a tripartite model of affect regulation. *Cognitive Therapy and Research*, *36*(6), 815–826.

Kleinbaum, D., Kupper, L., Nizam, A., & Rosenberg, E. (2013). *Applied regression analysis and other multivariable methods*. Nelson Education.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology. *Biological Psychiatry*, *44*(12), 1248–1263.

Luborsky, L. (1976). Helping alliance in psychotherapy. In J. L. Cleghhorn (Ed.), *Successful psychotherapy* (pp. 92-116). New York: Brunner/Mazel.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335–343.

Mallinckrodt, B. (2010). The psychotherapy relationship as attachment: Evidence and implications. *Journal of Social and Personal Relationships*, *27*(2), 262–270.

Mallinckrodt, B., & Jeong, and. (2014). Meta-analysis of client attachment to therapist: associations with working alliance and client pretherapy attachment. *Psychotherapy Theory Research & Practice*, *52*.

Marta-Simões, J., & Ferreira, C. (2018). Self-to-others and self-to-self relationships: Paths to understanding the valence of body image and eating attitudes in emerging adult women. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity*, 1-8

Matos, M., Gouveia, J. P., & Duarte, D. (2015). Constructing a self protected against shame: The importance of warmth and safeness memories and feelings on the association between shame memories and depression. *International Journal of Psychology and Psychological Therapy, 15,* 317–335.

Mehrabian, A. (1980). *Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies*. Oelgeschlager, Gunn & Hain.

Mendes, A. L., Ferreira, C., & Trindade, I. A. (2018). The central role of self-reassurance to explain body and eating attitudes. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity*.

Miller, J. B., & Hoicowitz, T. (2004). Attachment contexts of adolescent friendship and romance. *Journal of Adolescence*, *27*(2), 191–206.

Mitchell, R. L., & Phillips, L. H. (2007). The psychological, neurochemical and functional neuroanatomical mediators of the effects of positive and negative mood on executive functions. *Neuropsychologia*, *45*(4), 617–629.

Moreira, J., de Fátima Silva, M., Moleiro, C., Aguiar, P., Andrez, M., Bernardes, S., & Afonso, H. (2003). Perceived social support as an offshoot of attachment style. *Personality and Individual Differences*, *34,* 485–501.

Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, *23*(2), 242–266.

Neff, K. D. (2016). Does self-compassion entail reduced self-judgment, isolation, and over-identification? A Response to Muris, Otgaar, and Petrocchi (2016). *Mindfulness, 7*(3), 791–797.

Neff, K. D., & Vonk, R. (2009). Self-Compassion Versus Global Self-Esteem: Two Different Ways of Relating to Oneself. *Journal of Personality, 77*(1), 23–50.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York, NY, US: Oxford University Press.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30–80.

Panksepp, J. (2010). Affective neuroscience of the emotional BrainMind: Evolutionary perspectives and implications for understanding depression. *Dialogues in Clinical Neuroscience*, *12*(4), 533–545.

Panksepp, J., Fuchs, T., & Iacobucci, P. (2011). The basic neuroscience of emotional experiences in mammals: The case of subcortical FEAR circuitry and implications for clinical anxiety. *Applied Animal Behaviour Science*, *129*(1), 1–17.

Perris, C., Jacobsson, L., Lindström, H., von Knorring, L., & Perris, H. (1980). Development of a new inventory assessing memories of parental rearing behaviour. *Acta Psychiatrica Scandinavica*, *61*(4), 265–274.

Petrocchi, N., Pistella, J., Salvati, M., Carone, N., Laghi, F., & Baiocco, R. (2019). I embrace My LGB identity: Self-reassurance, social safeness, and the distinctive relevance of authenticity to well-being in italian lesbians, gay men, and bisexual people. *Sexuality Research and Social Policy*.

Pinto, C., Ferreira, C., Mendes, A. L., & Trindade, I. A. (2017). Social safeness and disordered eating: Exploring underlying mechanisms of body appreciation and inflexible eating. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, *22*(2), 303–309.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(3), 715–734.

Priel, B., & Shamai, D. (1995). Attachment style and perceived social support: Effects on affect regulation. *Personality and Individual Differences*, *19*(2), 235–241.

Rice, K. G., Ashby, J. S., & Slaney, R. B. (1998). Self-esteem as a mediator between perfectionism and depression: A structural equations analysis. *Journal of Counseling Psychology*, *45*(3), 304.

Richter, A., Gilbert, P., & McEwan, K. (2009). Development of an early memories of warmth and safeness scale and its relationship to psychopathology. *Psychology and Psychotherapy: Theory, Research and Practice*, *82*(2), 171–184.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, *21*(2), 95.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(39), 1161–1178.

Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, *125*(1), 3.

Russell, J. A. & Carroll, J. M. (1999). The phoenix of bipolarity: A reply to Watson and Tellegen (1999). *Psychological Bulletin, 125*(5), 611-617

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78.

Sarason, B. R., Pierce, G. R., Shearin, E. N., Sarason, I. G., Waltz, J. A., & Poppe, L. (1991). Perceived social support and working models of self and actual others. *Journal of Personality and Social Psychology, 60*(2), 273.

Silva, C., Ferreira, C., Mendes, A. L., & Marta-Simões, J. (2019). The relation of early positive emotional memories to women's social safeness: The role of shame and fear of receiving compassion. *Women & Health, 59*(4), 420–432.

Sroufe, L. A., & Waters, E. (1977). Attachment as an organizational construct. *Child Development, 48*(4), 1184–1199.

Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In *Anxiety and the anxiety disorders* (pp. 681–706). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science, 10*(4), 297–303.

Thayer, R. E. (1997). *The origin of everyday moods: Managing energy, tension, and stress*. Oxford University Press, USA.

Tomkins, S. S. (1984). Affect theory. *Approaches to Emotion, 163*(163–195).

van Bel, D. T., Smolders, K., IJsselsteijn, W. A., & de Kort, Y. (2009). Social connectedness: Concept and measurement. *Intelligent Environments, 2*, 67–74.

Watson, D., & Clark, L. A. (1985). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96*(3), 465.

Watson, D., Clark, L. A., McIntyre, C. W., & Hamaker, S. (1993). Affect, personality, and social activity. *Journal of Personality and Social Psychology, 63*(6), 1011.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Watson, D., Gámez, W., & Simms, L. (2005). Basic dimensions of temperament and their relation to anxiety and depression: A symptom-based perspective. *Journal of Research in Personality*, *39*, 46–66.

Watson, D., & Naragon, K. (2009). Positive affectivity: The disposition to experience emotional states. In *The Oxford handbook of positive psychology* (pp. 207-15). New York, NY: Oxford University Press

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*(2), 219–235.

Watson, D., & Tellegen, A. (1999). Issues in dimensional structure of affect—Effects of descriptors, measurement error, and response formats: Comment on Russell and Carroll (1999). *Psychological Bulletin*, *125*(5), 601–610.

Watson, D., & Walker, L. M. (19960101). The long-term stability and predictive validity of trait measures of affect. *Journal of Personality and Social Psychology*, *70*(3), 567.

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology*, *43*(1), 111–122.

Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The multidimensional scale of perceived social support. *Journal of Personality Assessment*, *52*(1), 30–41.