Narcissistic aggression: lab-based and naturalistic approaches to establishing subtype differences

William Smith

Supervisors: Dr. Claire Lawrence, Dr. Ruth Filik, Dr. Claudia Danielmeier

A thesis submitted to the University of Nottingham
for the degree of PhD in Psychology.

April 2022

Acknowledgements

I would like to thank my initial supervisor, Dr. Claire Lawrence, whose guidance and faith in me helped me to develop this project from a rough idea in an undergraduate email to a two-hundred page doctoral thesis. I am profoundly grateful to her for the opportunity she afforded me. I would also like to thank Dr. Ruth Filik and Dr. Claudia Danielmeier, whose specialist guidance as secondary supervisors helped to keep the project on track as it explored novel methods, and whose help in the months preceding the submission of this thesis went above-and-beyond what could be expected. I extend particular thanks to Dr. Filik for assuming the role of primary supervisor and providing extensive support despite the transition occurring at a challenging and late stage of the project.

I sincerely thank the hundreds of participants who took part in the research documented here across the three years of the project, and to my friends and colleagues from the School of Psychology who provided an eccentric and supportive community.

Table of Contents

Thesis Overview

The thesis is structured as a comprehensive theoretical chapter followed by a series of empirical chapters, which document several studies exploring the concept of narcissistic aggression in terms of two broad questions; what are the defining behavioural and cognitive characteristics of narcissistic aggression, and how are these characteristics influenced by variation in both subtype traits (individual differences within trait narcissism) and social context. The theoretical chapter should be considered a general introduction to the thesis, setting out the justification for studying the association between narcissism and aggression, along with the theoretical and literary background of the thesis, and a formal statement of a framework describing narcissistic aggression that forms the theoretical backbone of the subsequent empirical chapters that constitute the main body of the thesis.

The framework as it is described in the theoretical chapter presents a temporal description of narcissistic aggression as a process that evolves through a semi-linear sequence of cognitive and behavioural processes. This description is used to clearly situate the relationship between these processes and their place within the overall psychological 'narrative' of narcissistic aggression. However, in the following empirical chapters, this temporal description of narcissistic aggression is segmented into a sequence of distinct cognitive and behavioural processes, that are examined in detail independently of one another. These processes, and the order in which they are investigated, are threat-related behaviour, threat-related cognition and processes associated with fundamental cognition adjacent to threat-related cognition. These processes are operationalised using appropriate but exploratory lab-based measures, which are a behavioural task gauging response to social exclusion, threat-related vignettes, and a psychophysiological pupillometric measure respectively. Finally, threat-related behaviour is revisited in the final chapter as linguistic behaviour in a naturalistic social context, situated in contrast to the lab-based context of the preceding investigations.

Each empirical chapter therefore explores the two primary questions outlined above in the context of a different aspect of the theoretical framework. That is to say, the results documented in each chapter are examined both in terms of how they relate to the expectations of the relevant aspect of the theoretical framework, and how the results reflect the influence of differences in narcissistic subtype traits. This overview section is intended to provide a concise summary of each chapter situated in an overall narrative.

Chapter 1. In this literature review chapter, theoretical descriptions of aggression, narcissism and narcissistic aggression are explored in detail. Whilst superficially straightforward, an aim of this chapter is to illustrate that these traits become complex and multifaceted when scrutinized. A necessarily concise summary of the most important literature is presented, with reference to the different attempts to operationalize and measure each trait, along with reflections on the methodological challenges and concerns presented by some of these attempts. Finally, in order to establish theory and terminology that will be referenced throughout the thesis, the chapter also presents a specific definition of aggression, narcissism and narcissistic aggression, and assumptions regarding how these traits manifest in behaviour. The chapter concludes with the statement of a more formal theoretical description of narcissistic aggression, which serves as a theoretical framework underpinning the design and interpretation of the following empirical chapters, where elements of this framework are investigated in greater detail.

Chapter 2. The first empirical chapter establishes a starting point for the following studies, providing a straightforward empirical account of narcissistic aggression, investigating narcissistic threat-related behaviour in response to social rejection. Given the importance of understanding subtype differences in narcissistic aggression, the study documented by this chapter attempts to unify the somewhat diverse accounts of narcissistic aggression in the literature by investigating the influence of narcissistic subtype traits in response to the same threatening situation (i.e. social rejection), providing a necessary account of subtype differences to facilitate the

design and interpretation of later chapters. Finally, to respond proactively to the methodological concerns highlighted in Chapter 1, this chapter begins the important process of validating potentially useful measures of narcissism, aggression and related traits for use in the following empirical chapters.

Chapter 3. The following study expands on the results documented in the first empirical chapter by investigating the influence of narcissism and narcissistic subtypes on threat perception, in order to detect effects that may occur adjacent to threat-related behaviour whilst not manifesting behaviourally. This is particularly relevant in elucidating the influence of vulnerable narcissism, which might manifest more strongly in cognitive processes upstream of behaviour. This study also investigates narcissistic aggression in response to threats that are ambiguous in nature, in contrast to the relatively unambiguously threatening content of the situation featured in Chapter 2. Ambiguity may allow individual differences to be more clearly identified, as whether or not an ambiguous situation is interpreted as threatening is determined more by perception and less by the objective features of that situation.

Chapter 4. The third empirical chapter documents a deeper exploration of processes of threat perception in narcissistic aggression by investigating whether threat manifests more fundamentally as a heightened sensitivity to detecting conflicts between expected and observed outcomes. This description, driven by recent neuroscientific insights, potentially offers a more precise account of the cognitive processes through which narcissism influences aggressive behaviour. Furthermore, this description raises the possibility of reducing the operationalisation of narcissistic aggression to more objective assessments, based on physiological (pupillary response) and behavioural (flanker task) measures of conflict detection. The two studies documented in this chapter present an investigation of the link between narcissism and conflict detection using two different tasks; a ‘pure’ measure of conflict detection in terms of a letter-based flanker task, and a more ‘socially-relevant’ measure of conflict detection in terms of a facial emotion-based flanker task. Together, these tasks clarify the extent to which this proposed relationship

between narcissistic aggression and conflict detection is robust, and how important the meditating influence of socially-relevant stimuli is in this relationship.

Chapter 5. The fourth empirical chapter pivots from the approach taken in the previous chapters, documenting an exploratory study that moves away from lab-based operationalisations of narcissistic aggression in favour of using a naturalistic dataset and a mixed methods approach. The study utilizes webscraping to extract a large volume of text data consisting of contributions from individual members of a community on the major social media website Reddit, which is analyzed using both a qualitative thematic analysis and a quantitative text frequency analysis. This pivot offers an alternative perspective, viewing narcissism in naturalistic and social linguistic behaviours, without the potentially occluding influence of operationalisations of narcissistic aggression relying on self-report or artificial behavioural tasks. This shift in perspective may provide insights into how narcissistic behaviour manifests more organically when influenced by forces that are less present in a laboratory context. Adopting this approach, the study documented in this chapter explores both the potential utility of studying narcissistic aggression as a qualitative phenomena, and the more general utility of studying narcissistic behaviour in naturalistic datasets. The mixed methods aspect of this study builds on the existing literature examining narcissism in naturalistic online datasets with text frequency analysis, attempting to replicate and expand upon these findings by investigating the manifestation of narcissism and narcissistic subtypes in text frequency data. Importantly, this is made possible by taking a synergistic approach where qualitative insights guide the interpretation of text frequency data.

Chapter 6. The thesis concludes with a concise summary of findings and their implications, both in terms of their potential contributions to the study of narcissistic aggression and the practical and theoretical issues they highlight, and finally a closing reflective postface.

1. A theoretical review of narcissistic aggression

1.1. Aggression

1.1.i. Features of aggression

Narcissistic aggression occupies a space within a larger family of possible forms of aggression, that is defined by certain shared features, factors and rules. The unique description of narcissistic aggression emerges from a particular way in which these general characteristics of aggression interact, and therefore in order to understand what is meant by narcissistic aggression, it is necessary to first explore what is meant by aggression. There are many ways to approach a definition of aggression. The approach taken in the thesis is rooted in cognitive psychology and the study of individual differences. That is to say, aggression is understood as a cognitive-behavioural phenomenon that can manifest in all individuals independently of a specific cultural, historical or clinical context, that can be measured in a laboratory, and that can in some way be thought of as consistent in its manifestation between individuals or within an individual over time. In order to further develop this understanding of aggression, it is necessary to first explore and affirm definite features that describe how aggression manifests and is maintained as a cognitive-behavioural phenomenon, and then plot out how personal and situational factors intersect with these features to describe how narcissistic aggression emerges as a distinct phenomenon.

External features

Aggression has several external and superficially obvious features that manifest in behaviour. These features describe the behavioural component of a cognitive-behavioural approach to defining aggression. Considering aggression as a process with an initial triggering phase, a processing phase, and then a final outcome phase (DeWall & Anderson, 2011), behavioural features are most often situated at the end

in the outcome phase, representing whatever action was deemed to be necessary during the processing phase given the content of the triggering phase. However, it is important to remember that aggression is not necessarily contained to a single instance of this process, and outcomes can feedforward to trigger future instances of aggression (Anderson & Bushman, 2002; Miller et al., 2003), or can even be exploratory actions taken as part of the processing phase. This will be explored in more detail below, but for now it suffices to say that aggressive behaviour is not only an end but can also be the means to an end. The simplest feature-based definition of aggressive behaviour is that it can be either physical, as in interacting aggressively with the environment or another individual directly using physical force, or verbal, as in communicating to another individual a message containing aggressive content (Liu, 2004). This distinction highlights that even at the most basic level, there are different ways an individual might satisfy an aggressive intent through behaviour. In practice, there are many different forms of physical and verbal aggression, and not all are meaningful to compare or group together. For example, verbal aggression could be a negative description of someone, it could be a deliberately misleading comment, it could be a direct insult, and so on. Likewise, physical aggression could be breaking something, slamming a door, stealing or concealing something, or more directly physical actions such as assuming a threatening gesture, posture, or trying to intimidate. Regardless, whilst diverse, these all constitute possible superficial behavioural features of aggression.

As well as these superficial features of aggressive behaviour, there are also behavioural features that are concerned less with describing behaviour in terms of its basic characteristics, and concerned more with describing behaviour in terms of its evolution along a timeline. In general, aggressive behaviour follows a timeline with either reactive or proactive features (Dodge, 1991). Proactive features describe forms of aggressive behaviour that evolve over a longer period of time, and follow a pre-existing or otherwise structured script describing a sequence of actions. Furthermore, aggressive behaviour is proactive when the behaviour is deliberate and executed by an individual with a premeditated justification and outcome (Vitaro & Brendgen, 2005). For example, an individual identifies themselves as the receiver of some form

of unfair treatment or treatment that otherwise justifies an aggressive response (henceforth referred to as an 'insult', and what this might constitute is explored below). They then plan an aggressive response appropriate given the parameters of the insult, in order to achieve their desired outcome. This outcome might be influencing another individual's behaviour in order to reduce the future likelihood of the insult recurring, it may be punishing another individual, it may be simply achieving a feeling of catharsis reward, or it may be any combination of the above. This outcome might be achieved directly by directing aggression at the perpetrator of the insult, or indirectly by targeting individuals connected to the perpetrator in some way. It is important to appreciate the complex and layered nature of this process of planning an aggressive response to achieve a desired outcome, as the social environment in which insults occur and in which proactive aggression plots its course is not so simple in practice. That is to say, proactive aggression will probably follow a more abstract script than simply identifying an insult and then taking actions to punish the perpetrator. This is especially in the case of narcissistic aggression, in which proactive aggression can emerge through ruminating over ambiguous perceived insults (Okada, 2010), where the process of interpreting the insulting content and planning an appropriate behavioural response are likely to be equally ambiguous and not straightforward.

In contrast, reactive features describe forms of aggressive behaviour that evolve spontaneously without any deliberate planning (Dodge, 1991; Vitaro & Brendgen, 2005), either as an automatic instinctive response to an insult, or as a deliberate but fast response with little to no planning, driven instead by rapidly evolving and transient situational forces produced by the insult. As a result, whilst reactive aggressive behaviour also features a justification, a process of selecting an appropriate action, and performing this action to achieve a desired outcome, the nature of each of these features is different. The justification is wholly embedded in the content of the insult, the process of action selection is fast or automatic, and the desired outcome is likely more directly or immediately related to the content of the insult. For example, an individual may experience acute feelings of threat caused by an insult, that require or otherwise precipitate an immediate defensive behavioural

response. This likely takes the form of some direct aggressive action to take control of the situation, and remove the perceived source of the insult causing the threat. Again, it is important to appreciate the textured nature of reactive aggressive behaviour, as whilst these actions likely most often take the form of retaliation against the perceived perpetrator of the insult, they may also take the form of cathartic aggressive behaviour directed elsewhere, particularly in the case of narcissistic aggression where indirect reactive behaviour may be favoured when direct reactive behaviour incurs too many social costs.

Internal features

Aggression also has internal features that are not as obvious as behaviour, and provide the background mental context within which aggressive behaviour is shaped. Describing aggressive cognition is critically important, as it is through these features that narcissism as a personal factor interacts with aggression to assume the distinct form of narcissistic aggression. Internal features describe the cognitive component of a cognitive-behavioural definition of aggression, and are situated at the initial triggering phase and the processing phase of the process of aggression described above, and so interpret the content of the insult and select an appropriate behavioural response, if any. Furthermore, as well as acting as a precursor to behaviour, cognition arguably provides a broader and more accessible (if perhaps less immediately satisfying) set of outcomes, including rumination, hostile thought or visualisation, a negative change in attitude towards- or perception of the perpetrator.

The features of aggressive cognition can be described in terms of knowledge structures and cognitive priming and preparedness. Knowledge structures underlie all human cognition and guide the interactions between observations, expectations, cognitive and behavioural responses, and the updating of the same knowledge structures in real time based on feedback from these interactions (Anderson & Bushman, 2002). In aggressive cognition, knowledge structures provide scripts that describe the sequence of steps following an insult, such as how to interpret an insult and how to respond accordingly (Anderson & Bushman, 2002; DeWall & Anderson,

2011). As a feature of aggression, knowledge structures describe aggressive cognition as a structured process that follows a script. For example, an aggressive knowledge structure relating to how to respond when ignored by a friend on social media might direct an individual to interpret this as an act of social rejection, which should be accompanied by an emotional response of anger representing feelings of betrayal, and direct the individual towards any number of appropriate cognitive or behavioural outcomes that allow the individual to express their aggression, such as cognitively reframing the friend as an someone disliked, sending a verbally aggressive message, rescinding any current or future support, and so on. The knowledge structure allows the individual to access and navigate aggression, and gives them a sense of both the justification of their actions and an expectation of the outcome.

In practice, knowledge structures underlying aggressive cognition may be more or less complex than this. For example, structures underlying reactive aggressive responses may be very simple, and more akin to a reflex in their content, being little more than a script linking an experience with an automatic response. In contrast, knowledge structures underlying the longer and more deliberate proactive aggressive responses may be less easily mapped, being more akin to a complex web of related scripts guiding the individual through a large space of potential aggressive actions. Knowledge structures are persistent between episodes of aggression, however they are not inflexible, and can be updated with new information when the content of a knowledge structure ceases to be useful for navigating an episode of aggression (DeWall & Anderson, 2011). Nor are they isolated, as in practice any given individual knowledge structure will need to reference a diverse range of related information from other knowledge structures. For example, the script dictating an aggressive response to rejection from a friend on social media necessarily references scripts relating to social media and platonic relationships more generally, and inevitably also references more abstract self-referential knowledge structures. This observation is crucial to understanding the features of narcissistic aggression, which does not necessarily have anything to do with unique knowledge structures of aggressive cognition, but has everything to do with how related knowledge structures concerning social behaviour and self-perception influence aggressive cognition.

Where knowledge structures describe the features of the framework of aggressive cognition, priming and preparedness describe the features determining how sensitive this framework is to activation, or, in other words, determining the scope of potential situations in which aggressive cognition is deemed appropriate (DeWall & Anderson, 2011; Huesmann, 1998). Priming refers to the sensitivity of aggressive cognition to state-dependent forces (Engelhardt & Bartholow, 2013), which may be characteristics of the situation such as current or recent experiences, transient personal characteristics such as mood, or an interaction of both in the form of unique personal interpretations of otherwise ambivalent situational characteristics. These forces calibrate the general cognitive state of the individual to make them more or less inclined towards aggressive cognition. For example, an individual in an uncomfortable environment and a negative mood may experience alterations to their general cognitive state that make them more likely to interpret an ambiguous event as negative or insulting, and more sensitive to unambiguous insults. As a result, they become more likely to initiate an aggressive episode in response to an event that would in other conditions not be considered sufficient justification for aggression. In other words, these alterations to their general cognitive state have primed them to respond with aggression. This feature provides an important description of the state-dependent nature of aggressive cognition. Preparedness describes the baseline sensitivity of aggressive cognition in response to an event that might potentially trigger an aggressive episode (DeWall & Anderson, 2011). This feature is not state-dependent, but rather represents the influence of more static characteristics of an individual, specifically state-independent personal characteristics such as personality and prior beliefs. For example, an individual with certain attitudes that justify aggression in response to a broad range of situations is more likely (more "prepared") to engage in aggressive cognition simply by virtue (or perhaps vice) of their nature.

This is how knowledge structures not directly related to describing an aggressive episode influence aggressive cognition, by setting the threshold for what does and does not justify an aggressive response, based on more general perceptions and beliefs (Slotter & Finkel, 2011). More specifically, and most importantly, this is how knowledge structures shaped by aggression interface with aggressive cognition. In

this sense, narcissism is arguably a state-independent personal characteristic that influences aggressive cognition through altering preparedness.

Summary

Taken together, behavioural features of aggression describe the actions that constitute aggression, and the more general reactive or proactive sequences within which these actions occur. Cognitive features describe the framework of knowledge guides cognition and underlies these behavioural sequences, and also describe the sensitivity of aggressive cognition to state-dependent forces, and the baseline predisposition to aggressive cognition given the state-independent characteristics of an individual. This feature-based description of aggression (Anderson & Bushman, 2002; DeWall & Anderson, 2011) is important to the exploration of narcissistic aggression in this thesis, as it sets out the basic cognitive-behavioural architecture of aggression. However, throughout this description of features, there has been frequent reference to situational and personal characteristics that provide the external and internal context respectively to an aggressive episode. These characteristics can be considered independently of the features they interact with, and in doing so can lay the foundation for a less rigid definition of aggression, that references features whilst not being wholly described by them.

These characteristics are referred to as risk factors of aggression (i.e. personal or situational risk factors) - traits and states that act as forces mediating the features of aggressive behaviour and cognition. These factors should not be considered auxiliary, but rather the presence of these factors constitutes a fundamentally important condition of aggression, that gives aggression its diverse and dynamic nature. Indeed, it is narcissism as a personal factor that gives narcissistic aggression its unique qualities, and so examining these risk factors is a central focus of the thesis. In the next section of this chapter, a holistic approach to combining both features and factors is explored, leading to a picture of aggression as a multidimensional cognitive-behavioural phenomenon that is not described as a neatly delienated set of features, but rather as coordinates along a set of cognitive and behavioural axes that

have different but not mutually exclusive features. This sets the stage for defining and exploring narcissistic aggression, which is better understood as a multidimensional trait due to the subtle and apparently inconsistent influence of subtype-level narcissistic traits (this is explored further in **Chapter 1.2. Narcissism**).

1.1.ii. Theories of aggression

There are several important components to the theoretical approach to describing aggression used in this thesis, with each component providing a particular account of how the static features of aggression interact with one another and are influenced by personal and situational factors to create a dynamic cognitive-behavioural process.

The General Aggression Model

The General Aggression Model (GAM; Allen et al., 2018; DeWall & Anderson, 2011) describes how cognitive factors influence and direct the way aggression unfolds within the context of a given individual and a given situation. Broadly, the GAM focuses on how personal and situational inputs are interpreted through the lens of an individual's enduring knowledge structures and transient cognitive and emotional states. Furthermore, the GAM emphasises that the type of cognitive appraisal used by the individual to perform this task of interpretation can strongly influence the outcome. The GAM places personal and situational factors as the proximate causes of aggression, in the sense that they are situated upstream of all ensuing cognitive processes. Personal and situational factors interact to facilitate or inhibit one another. For example, narcissism alone is unlikely to be a sufficient cause of aggressive behaviour, but rather narcissism is facilitated by the presence of certain situational factors. These factors then activate a set of internal states which influence the likelihood of aggressive behaviour occurring. For example, when socially rejected (situational factor), a narcissistic individual (personal factor) may experience a state of negative emotional arousal and rumination (transient internal states) and may recall hostile beliefs and perceptions of others that are implicit consequences of narcissism

(enduring internal states). This combination of internal states is then evaluated by processes of cognitive appraisal that can be either automatic or controlled. Automatic appraisal tends to occur immediately, whereas controlled appraisal is slower and functions as an executive regulating the outcome of automatic appraisal. The outcome of these appraisal processes is to select an appropriate strategy in response to these internal states, which may also be to make no response (DeWall & Anderson, 2011).

Repeated successful use of a strategy (i.e. positive feedback from the outcome of appraisal) results in a feedback loop wherein future appraisals involving that strategy shift towards automation. When internal states of negative arousal linked to narcissistic cognition are successfully controlled using aggression, this process of feedback weighted towards automation underlies the development of reactive aggressive tendencies (i.e. after proactive forms of aggression result in positive outcomes for a narcissistic individual). This is of particular relevance to the development of narcissistic aggression outside of the cycle of a single episode of aggression. The GAM also highlights that this feedback loop does not influence only the individual, but also influences the social context in which the individual is behaving. Aggression occurs in a cycle of escalation, wherein aggressive behaviour provides a successful short-term strategy, but in the long-term serves to perpetuate and even amplify the situational conditions that trigger aggression. It is through this cycle that narcissism produces a self-fulfilling prophecy where perceived hostility is justified by the conditions that it creates.

This theory provides a description of how narcissism can interact with situational forces to produce aggressive behaviour, and how this behaviour can be maintained through becoming an ingrained feature of the narcissistic individual's internal and situational context. An alternative way of representing this balance of situational forces, personal traits and executive control (i.e. appraisals) is in terms of thresholds (Slotter & Finkel, 2011). This is a theoretical description that is useful in an experimental psychology approach as it allows the process of aggression to be modelled as following rules whilst remaining continuous and sensitive to the shifting

influence of factors. In this description, it is assumed that when necessary conditions are satisfied and an aggressive action becomes possible (i.e. situational or personal triggers), there is a defined threshold at which an aggressive action is deemed appropriate. There is also an overall measure of cognitive and/or affective arousal that lies either below or above this threshold, and therefore ultimately determines whether an aggressive action is made. The overall state of arousal is the product of a combination of 'exciting' and 'inhibiting' factors that increase or decrease arousal respectively, and vary depending on the person and situation (for a more detailed theoretical account, see I^3 theory; Finkel, 2007).

The personal and situational factors described by the GAM are considered as inputs, which can be moderated by internal states to give an overall state of arousal that is considered relative to a threshold by processes of appraisal. This quantitative and easily modelled description of aggression informed by the theoretical predictions of the GAM will be used as the foundational theoretical description of aggression in this thesis.

A behavioural systems account of aggression

Whilst this provides an account of when and why a narcissistic individual may engage in aggression in terms of transient and enduring features of the individual, it is also useful to consider a theoretical description of why these features should be associated with aggression at all. That is to say, why a given individual should be motivated to resort to aggressive behaviour in the first place. This will be explored with specific reference to narcissism more thoroughly in the later in this chapter (see **Chapter 1.2. Narcissism**), whereas the description here will focus more generally on the motivational component of aggressive cognition. Behavioural systems theory (Bowlby, 1982) situates behaviour and cognition in terms of a motivational system, in which behaviour and cognition are strategies used to either approach or avoid the primary goal of that system. As many goals are not immediately achievable or otherwise challenging to achieve, behavioural systems have contingency strategies

used to overcome barriers to the achievement of goals; hyperactivating and deactivating strategies. Hyperactivating strategies intensify the pursuit of the primary goal of the system, to ensure the focus of the system remains goal-oriented despite challenges until the goal is achieved. However, in instances where the goal is deemed to be too difficult to achieve, or where repeated use of hyperactivating strategies has failed to result in the attainment of the goal, the system can deploy deactivating strategies to intensify goal-avoidant cognition and behaviours, and thereby mitigate the emotional and physical toll of further goal pursuit.

Both of these strategies therefore have an obvious adaptive function. However, hyperactivating and deactivating strategies also present obvious mechanisms for the development of maladaptive motivational forces underlying aggression (Shaver et al., 2011). This is because the extent to which these strategies are used, and the sensitivity of their threshold for activation (i.e. when goal pursuit is sufficiently challenged for these strategies to be appropriate), are flexible and can be influenced by personal and situational factors through similar mechanisms as described above. That is to say, some individuals are motivated towards using aggression to achieve their goals by a reduced tolerance for barriers to goal pursuit when using otherwise non-aggressive strategies. In the context of aggression, hyperactivation represents the engagement of assertive aggressive actions to 'brute force' or coercively pursue a goal. Deactivation, on the contrary, does not result in the engagement of assertive aggressive actions, but nevertheless introduces a heightened sensitivity to threat. This is because successful use of deactivating strategies in ambiguous social situations does nothing to challenge (and indeed reaffirms) the perception that these situations are alarming and that the individual has avoided a potential threat. This works to maintain the negative cognitive biases that underlie hostility. This hostility, in turn, precipitates later episodes of aggressive behaviour and expands the scope of situations in which an individual might resort to aggressive hyperactivating or deactivating strategies.

Insofar as the ability to manage the pursuit of goals and adjust for changing circumstances is a basic feature of human cognition, and motivation provides a guiding force through which appropriate strategies are employed, an individual may feel motivated to use aggression because hyperactivating and deactivating strategies featuring aggression are simply more accessible parts of their toolkit of goal pursuit strategies than might otherwise be typical (Huesmann et al., 2003; Huesmann & Kirwil, 2007). This is both a trait property of the individual, and a property that is facilitated and maintained through repeated use of these strategies (and in particular deactivating strategies that enhance hostile cognition). A higher-level description of this motivation situated in the beliefs and attitudes of the individual might manifest as a strong conviction in the utility and effectiveness of aggressive behaviour and cognition as problem-solving tools, but also a more covert aversion to failure that 'raises the stakes' to a sufficient extent that aggression is perceived as necessary. That is to say, aggressive motivation is characterised not only by a direct motivation to make use of aggressive behaviour and cognition, but also an indirect motivation to use aggression as a drastic but justified response to an urgent fear that the individual will fail to achieve their goal (Keltner et al., 2003). This fear could also be represented as a desire to justify sunk costs, for example if the individual perceives that they have worked too hard or have too much emotional investment in a goal to give up pursuing it. Simple economic considerations therefore persist even at this higher level of description, as aggressive motivation is primarily concerned with ensuring resources invested in pursuit of a goal are not wasted, and also establishes the conditions whereby taking drastic measures to prevent this waste is justified. Personal and situational factors may lower the threshold at which sufficient investment has been made in achieving a goal that aggressive strategies become appropriate.

Invested resources can be understood literally to represent physical, cognitive or emotional exertion, all of which may have obvious physical correlates. Economic considerations like these are likely more relevant in simplistic forms of aggressive behaviour (e.g. frustration) although are certainly also relevant considerations in more

complex forms of aggression (Wilkowski & Robinson, 2008). However, where aggression is embedded in an individual's social context, these resources can also represent power and social influence (Bond, 2004). This is particularly relevant in the concept of narcissistic aggression, where the fear of losing social status is a driving force behind aggressive behaviour (Bogart et al., 2004; Grapsas et al., 2020), and where aggression is justified both by the effort invested in maintaining social status, and the severe consequences associated with losing social status given a fragile sense of self-esteem. A more abstract and refined psychological description of these traits of narcissism is described below (see **Chapter 1.2. Narcissism**), but as above, narcissistic aggression can be reduced to simple economic considerations, as the sunk costs of narcissistic behaviour along with the exaggerated consequences of losing social resources lower the effective threshold at which aggressive behaviour is deemed justifiable. Subtype differences in narcissistic aggression emerge depending on the disposition of the individual towards hyperactivating or deactivating strategies.

Summary. In summary, the GAM provides an overview of how factors and features involved in aggression fit together to describe an episode of aggression. Thresholds allow a more structured account of this cycle and how episodes of aggression resolve. Finally, a behavioural systems approach provides a description of the internal motivational forces that drive aggression, both as a high level psychological description and in basic economic terms.

1.1.iii. Studying aggression

There are several intuitive and obvious challenges to operationalising aggressive behaviour is notoriously difficult to operationalize. The first is that for the reasons described above, aggressive behaviour is not produced arbitrarily and cannot be elicited in any given individual in any given context. Secondly, aggression is not binary in degree or nature: there is no non-arbitrary threshold at which a behaviour can be classified as aggressive, and aggressive behaviour can manifest in multiple different ways, which may be aggressive or non-aggressive depending both on the

intentions and motivations of the participant, which are difficult to assess, and the theoretical perspective of the researcher, which will necessarily be reductive.

Nevertheless, decades of research has developed various methods for eliciting and operationalising aggressive behaviour in a laboratory setting. An exhaustive list is beyond the scope of this thesis. Broadly, methods of eliciting aggression include; physical stressors that prime an aggressive response, such as delivering electric shocks, loud noises or spiking food with unpleasantly intense spice (Bell & Baron, 1976; Chester & DeWall, 2016; Geen & Powers, 1971; Lieberman et al., 1999); embarrassing or annoying social situations such as being ignored or scolded (Ayduk et al., 2008; Cohen et al., 1996); competitive tasks involving directly working against other participants (Boccadoro et al., 2021), which may feature explicitly provocational or spiteful behaviour from rivals. Methods of operationalising aggression include; several psychometric scales that capture different aspects of aggressive cognition and behaviour; measures in behavioural tasks such as the severity of punishment delivered to another participant; physiological measures such as heart rate or skin conductance (McCarthy & Elson, 2018). As a result, there are a large variety of possible approaches to studying aggression, none of which will capture precisely the same phenomena. Throughout the thesis, care will be taken to an appropriate approach to operationalising aggression, with consideration to what type of aggression an approach is capable of studying, along with acknowledging its potential limitations.

1.2. Narcissism

1.2.i. Introduction to narcissism

Narcissism has a diverse history of study, and its conception as a psychological trait has not been consistent. This inconsistency has led to a popular understanding of narcissism that is poorly defined, and only relatively recently has a more cohesive academic understanding of narcissism become available. In this section, the

development of narcissism as a psychological trait will be described, in order to both illustrate its broad range of possible features and clearly delineate what does and not does constitute the current psychological characterisation of narcissism.

Early theorists (e.g. Freud, Rank; see Levy et al., 2011 for a comprehensive review of these and other early theorists described in this section) developed theories of narcissism that remain influential, specifically the concept of narcissism as a trait, and the idea that a narcissistic individual does not occur in an solipsistic 'vacuum' but rather requires a community of individuals to maintain them, and that the personal and situational features of this community might be influential. Wälder introduced more formally the concept of narcissism as a personality type with a set of defined characteristics, including a sense of superiority and egocentric self-focus, but also notably introducing lack of empathy as a characteristic. Importantly, these early theorists also developed influential ideas concerning the motivations underlying the narcissistic personality. Rank and Horney suggested that a narcissistic individual is not solely motivated by feelings of reward associated with their behaviour (e.g. the satisfaction of achieving attention, status, respect and so on) - in other words, not merely narcissism for the sake of narcissism - but is also motivated by a 'false' self-conception that must be actively defended and maintained. Actively supporting this narcissistic self-conception is necessary because it has no legitimate foundation that would otherwise offer passive support (i.e. as would a 'normal' self-conception). Feeling a sense of vulnerability considering self-concept or identity more generally is a naturally threatening state, and all individuals have a basic desire to maintain a secure sense of identity. In non-narcissistic individuals, there is a coherence between top-down expectations and bottom-up observations about that identity. That is to say, the ideas an individual has about their identity are largely coherent with what their observations tell them about the actual nature of their identity. However, the narcissistic individual experiences a decoherence, as warped narcissistic expectations regarding their identity are not confirmed by observations. As a result, they are motivated to actively use a suite of narcissistic behavioural and cognitive strategies to continually try and correct this state of decoherence. Classically, it is the grandiosity

of the narcissistic self-concept that creates this decoherence (Winnicott, 1984), and strategies to address this typically take the form of directly and forcefully reshaping the features of their social landscape to better fit grandiose expectations.

Early psychoanalytical theory helps to illustrate this defensive motivational aspect of narcissism. Freud classically suggested that narcissism was characterised by dynamic behaviours that served to avoid and repress the unconscious trauma that manifests when the narcissistic self is challenged. Reich suggested a psycho-developmental framework for narcissism wherein early traumatic experiences result in difficulties maintaining a stable sense of self-esteem. The resulting stress causes narcissistic individuals to retreat into a constructed grandiose inner world, where their self-esteem is not only stable but justified by a range of positive (i.e. grandiose) features such as superiority, popularity, strength, attractiveness and so on. However, as this self-conception is essentially artificial it is regularly challenged by ordinary experience, causing erratic oscillations in self-esteem that drive narcissistic cognition and behaviour. This frequently polarising self-concept contributes to the narcissistic individual developing a similarly polarised attitude towards themselves and the world around them, in which ambiguities are lost, and the only possible outcomes are either failure or perfection. As a result, narcissists may regularly experience feelings of shame or persecution.

Importantly, these defensive motivations also lead to narcissistic aggression. Indeed, the typically antisocial nature of narcissistic cognition and behaviour may be attributable to the more specific qualities of the (classically grandiose) narcissistic self-concept. Narcissistic individuals view themselves as being more intelligent and extroverted than others, but view themselves as being admired rather than liked (Raskin et al., 1991). Similarly, narcissists classically view themselves as being agentic as opposed to being communal (although this conception is challenged by the emergence of communal narcissism which is explored at length later in the thesis; see Gebauer & Sedikides, 2018), and do not consider themselves to be more agreeable or moral than others (and may even view themselves as being less agreeable; Campbell

et al., 2002). This highlights that the narcissistic self-concept is not contingent on the affection of others, but rather is contingent on a kind of respect that should follow naturally from recognition of the impressive positive traits possessed by the narcissistic individual. As a result, prosocial behaviours designed to win affection are less useful than assertive behaviours designed to enforce admiration.

Other researchers have been careful to challenge this conception and suggest that narcissism as a suite of behaviours and traits designed to defend and maintain a fragile identity may not capture the entire psychological landscape of narcissism (Baumeister et al., 2000; Baumeister & Vohs, 2001). Grandiose narcissistic beliefs about the self are not necessarily intended as protection against the lows of low self-esteem but might rather be a means of chasing an 'addiction' to the highs of high self-esteem. That is to say, narcissism may in some individuals be driven by a more straightforward desire to maintain a state of high self-esteem, rather maintaining a facade of high self-esteem for the purpose of defending against a state of low self-esteem. A simple functionalist approach to resolving this issue is to suggest that these two distinct motivations are effectively equivalent given their similar manifestation. That is to say, maintaining a state of high self-esteem through a strong defensive aversion to low self-esteem should involve many or all of the same behaviours as assertively and directly maintaining high self-esteem, and therefore the two motivations can be considered functionally the same. It may therefore not be useful to become too preoccupied with delineating specific differences in the underlying cognitive and motivational content of narcissism.

1.2.ii. Narcissistic behaviour and cognition

Narcissistic behaviours are naturally highly context specific, and any behaviour could arguably be labelled a narcissistic behaviour if it is motivated by a desire to maintain a narcissistic self-concept (Grapsas et al., 2020). However, there are also naturally several behaviours which are more typical of narcissism than others, due to being

particularly overt or routine, or being particularly effective as a means of achieving the desired outcome of narcissistic motivations.

Narcissism is associated with increased effort to excel in tests of performance, which may be any task or activity in which there is a potential for directly signalling positive attributes (e.g. intelligence, popularity, physical fitness, proficiency etc.) relative to others (Wallace & Baumeister, 2002). However, as narcissism increases the valence of changes in self-esteem, the enhanced reward of signalling positive attributes through success must be weighed against the more severe risk of unintentionally signalling negative attributes through failure. As a result, not every task will be sufficiently rewarding to justify the investment of assertive narcissistic behaviours, only tasks that prove particularly challenging or interesting (i.e. challenges with a particularly high reward associated with success) are likely to be attractive (Roberts et al., 2010; Wallace & Baumeister, 2002). If participating cooperatively with others in a test of performance, a feature of narcissism may be perceiving shared accomplishments as having shared rewards (i.e. reduced value for the individual), and therefore unlikely to attract narcissistic behaviours.

Indeed, an adjacent form of narcissistic behaviour may be creating and engaging in spaces in which competition or otherwise individual performance is emphasised (Benson et al., 2019; Zeigler-Hill et al., 2018). Typical examples of these may be online communities or platforms which facilitate this type of individualistic signalling of positive attributes, where a high level of control over the content of what is being signalled results in high reward and relatively low risk. Narcissistic behaviours here may involve adopting a particularly provocative or self-enhancing identity, and maintaining this identity through creating content of a controversial or self-enhancing nature that serves to both attract attention and communicate whatever perceived qualities the individual is motivated to enhance. This highlights how the seemingly contradictory combination of egotistical self-focus and preoccupation with social context that defines narcissism theoretically works in practice, as narcissistic

behaviour involves both attracting an audience and ensuring that audience attends to and is receptive of the positive attributes being signalled (Morf & Rhodewalt, 2001).

This apparently benign social posturing represents what could be called the 'light side' of narcissism, which is reflected in the popular conception of narcissistic behaviour as typically vain and extraverted. However, importantly, it is this characteristic of narcissistic behaviour to seek situations in which the narcissistic individual can signal positive attributes whilst minimizing the risk of losing status (and therefore self-esteem) that drives the distinctively aggressive 'dark side' of narcissistic behaviour (Kampe et al., 2021), which serves to coerce competition away from the spotlight and defend against situations where there is a perceived threat of losing status (and perhaps turn these situations into an opportunity for gaining status, if aggressive behaviour is considered a positive quality to signal; Grapsas et al., 2020; Zeigler-Hill et al., 2018). This aggressive narcissistic behaviour can be proactive, in order to pre-emptively respond to a predicted threat or facilitate a better space for signalling positive attributes, or reactive, in response to a perceived loss of status (Fossati et al., 2010; Reidy et al., 2010). This reactive response may constitute an automatic retaliation against the pain or anticipated pain of entering a state of low self-esteem, or as a frustrated attempt to assert an anticipated state of high self-esteem. Narcissistic aggression will be explored further as a distinct phenomenon below (see **Chapter 1.3 Narcissistic Aggression**).

1.2.iii. Grandiose and vulnerable subtypes

It is possible to refer to narcissism in a broad sense as above, as a suite of cognitive and behavioural tools motivated by the desire to maintain a state of high self-esteem contingent on reinforcing narcissistic self-perceptions. This form of narcissism could be described as 'trait narcissism'. However, it is possible to provide a more precise description of narcissism by examining the variation within narcissistic traits at a subtype level. It is now widely accepted that there are at least two distinct subtypes of trait narcissism that are reliably observable using several different measures

(described in the following section). These two subtypes are referred to as ‘grandiose’ (or ‘overt’) and ‘vulnerable’ (or ‘covert’) narcissism respectively (Wink, 1991; Yakeley, 2018).

Grandiose narcissism is characterised by extroversion and obvious external signalling of perceived positive attributes through overt narcissistic behaviours (Dickinson & Pincus, 2003; Miller et al., 2011). Grandiose traits include sensation seeking, entitlement, arrogance, attention seeking, and indeed a particular disposition to expressed exaggerated self-importance and status (i.e. grandiosity). As a result, grandiose narcissism emphasises traits that allow an individual to more effectively engage assertively with their social context to facilitate the narcissistic display of positive attributes. In contrast, vulnerable narcissism is characterised by introversion and a quiet or even modest demeanour (Dickinson & Pincus, 2003; Kealy & Rasmussen, 2012; Miller et al., 2011). Vulnerable traits include behavioural inhibition, hypersensitivity to criticism, neuroticism and anxiety. Importantly, whilst superficially vulnerable narcissism seems far removed from the narcissistic motivations which are more clearly evident in grandiose narcissistic behaviour, vulnerable narcissism also emphasises maintaining a state of high self-esteem consistent with narcissistic self-perceptions. However, whilst in grandiose narcissism this is achieved through assertively signalling positive qualities and thereby directly eliminating threats to self-esteem, in vulnerable narcissism this is instead achieved through actively avoiding situations which may signal negative qualities and thereby indirectly eliminating potential threats to self-esteem. This aversion to the risk of losing status is what constitutes the ‘vulnerability’ of vulnerable narcissism (Szymczak et al., 2020).

Grandiose and vulnerable narcissism are therefore distinct constructs but possess a shared motivation that defines trait narcissism more generally (Jauk et al., 2017). This highlights the slightly misleading nature of the terminology of ‘grandiose’ narcissism, as it implies grandiosity is unique to this subtype, when in fact vulnerable narcissism is just as much motivated by grandiose self-beliefs, but this grandiosity manifests in a

less overtly grandiose fashion (i.e. through private beliefs maintained through avoidance). Indeed, just as both subtypes possess grandiose narcissistic self-perceptions, they also both possess vulnerabilities and a hypersensitivity to loss of self-esteem (Jauk et al., 2017; Kealy & Rasmussen, 2012). Despite making use of different behavioural strategies, grandiose narcissism will nevertheless lead to rumination, inhibition and neuroticism when extroverted narcissistic behaviours fail to signal positive attributes, or even lead to a loss in status. It is therefore not the case that grandiose and vulnerable narcissism should be considered wholly independent traits that emerge upon closer examination of an otherwise oversimplified trait narcissism. Instead grandiose and vulnerable narcissism should be viewed as two dimensions of an overall and non-redundant trait narcissism construct.

1.2.iv. Studying narcissism

The well-developed theoretical description of narcissism allows plenty of scope for the study of narcissism using qualitative methods. However, given the persistent experimental psychological interest in investigating narcissism as a trait, quantitative psychometric approaches to studying narcissism prove the most widely used. A variety of psychometric tools have been developed. The most widely used measure is the Narcissistic Personality Inventory (Raskin & Hall, 1979), which has undergone extensive stress-testing to support its construct validity due to its relatively early inception in the study of trait narcissism (Emmons, 1984). The NPI also exhibits correlations with other measures that are consistent with the theoretical description of narcissism, such as self-focused attention, power-seeking, defensive reactivity to self-esteem threats, and low intimacy and agreeableness. As a result, the NPI is largely appropriate as a metric for trait narcissism generally. However, there are methodological limitations to the NPI that makes it less appropriate for more thorough examination of narcissistic behaviour and cognition.

The primary issue with the NPI is that it attempts to operationalise narcissism generally whilst apparently only measuring a subset of features within the overall

scope of narcissism. More specifically, vulnerable features of narcissism as described in the previous section appear entirely absent from the NPI, leaving a significant blind spot in its operationalisation of narcissism (Ackerman et al., 2011). As a result, the NPI could arguably be said not to measure narcissism, but rather a trait that is circularly defined by high scores on the NPI (i.e. the construct and the scale are equivalent). However, it may be more balanced to instead suggest that the NPI does not measure narcissism generally, but rather more precisely measures the grandiose features of narcissism, being as it was developed during a time when subtype variations in narcissism were not widely acknowledged. This is supported by the finding that NPI scores are also correlated with high extraversion, social boldness, sociality and dominance (Miller & Maples, 2011). A similar critique could be made of the Narcissistic Admiration and Rivalry Questionnaire, which operationalises narcissism in terms of its core affective and motivational processes, but features an exclusively grandiose conception of narcissism (Back et al., 2013).

There have been efforts to develop psychometric scales that capture features of narcissism neglected by the NPI and similar questionnaires. The Hypersensitive Narcissism Scale (Hendin & Cheek, 1997) attempts to more explicitly measure the sensitivity to criticism that may be implicit to the grandiosity assessed by the NPI, and in particular may define the motivation behind the avoidance-based behavioural strategies that define vulnerable narcissism. However, whilst use of the HSNS might expand the scope of narcissistic features open to psychometric investigation, the HSNS nevertheless also stops short of measuring the full scope of narcissistic features described above by neglecting other aspects of narcissistic vulnerability. For example, whilst it is certainly the case that vulnerable narcissism is associated with hypersensitivity to potential threats to self-esteem, it is also defined by features such as neuroticism and anxiety. Arguably, it is the combination of hypersensitivity with these neurotic and inhibited features that distinguish vulnerable narcissism as a distinct subtype, as hypersensitivity is otherwise shared with grandiose narcissism and indeed with basic trait sensitivity more generally. That is to say, a high score on the HSNS may reflect individuals with high trait vulnerable narcissism, but it may

equally reflect high trait grandiose narcissism, or simply individuals who are more sensitive to being provoked or frustrated by perceived self-esteem injury. This becomes particularly problematic in contexts in which narcissistic aggression is of interest, as attempting to examine the relationship between aggressive behaviour and vulnerable narcissism as measured by the HSNS is likely to be skewed by the more obvious relationship between aggression and sensitivity to impulsive or reactive aggressive behaviour.

Developed more recently, The Five-Factor Narcissism Inventory (Glover et al., 2012) attempts to address the shortcomings of both the NPI and HSNS by unambiguously measuring both grandiose and vulnerable narcissism, using a set of items that more holistically represents the defining features of these subtypes, as well as identifying more general subtype measures that encompass both vulnerable and grandiose features. Although it lacks the rigorous testing and between-measures validation of the NPI, the FFNI seems especially promising in its potential to provide a measure with better construct validity than both the NPI and HSNS, whilst encompassing the scope of both these measures within a single shared scale.

1.3. Narcissistic aggression

Based on the theoretical description of aggression provided above, narcissistic aggression can be understood straightforwardly as the result of narcissism acting as a risk factor driving the cycle of aggression as described by the GAM. However, the theoretical description of narcissism highlights that the association between narcissism and aggression is not trivial but rather a fundamental product of narcissistic motivations, that drive the deliberate engagement of aggressive behaviour and cognition. In this section, ideas from the review presented above are synthesised into an original theoretical framework of narcissistic aggression.

1.3.i. Characteristics of narcissistic aggression

Narcissism is a suite of behavioural and cognitive tools that help an individual select strategies to defend a fragile (or narcissistic) sense of self-esteem (Kampe et al., 2021; Kim & Barry, 2021; Stucke & Sporer, 2002). The increased sensitivity to threat necessarily associated with this fragility creates a state of cognitive alert, as harm to self-esteem is both considered unacceptable and constantly anticipated. Whilst many situations may be unambiguously threatening or non-threatening, many more situations may be ambiguous in their content. Given the exaggerated consequences associated with harm to self-esteem, narcissism creates a set of biases and automatic assumptions that help resolve this ambiguity and reduce the risk posed by ambiguously threatening situations. These narcissistic biases are defensive in nature and result in increased hostility and threat sensitivity (Kampe et al., 2021). This is due firstly to narcissism featuring an inherent presupposition of being under threat, and due secondly to the practical consideration that when a threat is ambiguous, and the consequence of overlooking this ambiguous threat is severe, then the safest strategy is simply to assume a threat is present. In other words, for the narcissistic individual, "if in doubt, assume hostile intentions" is the most reliable strategy given the conditions underlying narcissism. Aggressive behaviour and cognition naturally follows from these narcissistic biases, as an effective and immediate defensive response to threats that is appropriate given the often ambiguous nature of threats to self-esteem and the severe consequences they pose.

1.3.ii. The cycle of narcissistic aggression

An episode of narcissistic aggression is initiated by an event that is interpreted as threatening given the narcissistic traits of the individual. These traits are a combination of specific narcissistic features, such as a heightened self-focus and an incongruent self-image, and the cognitive structures they influence, such as self-esteem, empathy, and cognitive biases. In other words, narcissistic aggression is initiated when an event is perceived as threatening given the narcissistic individual's

fragile sense of self-esteem. As a result, these events frequently involve social insults, but the scope of what constitutes a social insult is very broad given the range of events that may - directly or indirectly - undermine an individual's identity, perceived self-worth or social standing (Bushman & Baumeister, 1998). This activates transient states of negative emotional arousal and anger, which may immediately precipitate an episode of narcissistic aggression if they increase arousal above a threshold, and/or lead to rumination and aggressive priming which lowers that threshold and makes a future episode of narcissistic aggression more likely.

The final stage prior to the actual initiation of narcissistic aggressive behaviour or cognition is an executive check that appraises the justification for aggression after information about a potential threat has been cross-referenced with knowledge structures influenced by narcissistic traits. This process of appraisal is subject to similar narcissistic biases, as naturally narcissistic traits influence the motivational decision-making processes that inform appraisal (explored further below). Furthermore, narcissistic traits shift the processes of appraisal involved in narcissistic aggression towards increased automation, given the urgent and high risk nature of potential threats to self-esteem.

1.3.iii. Narcissistic aggression as a behavioural system

Narcissism can be modelled as a behavioural system with the defence of self-coherence as its primary goal. That is to say, the fundamental threat posed by damage to a fragile (narcissistic) sense of self-esteem is the loss of self-coherence and the emotional trauma caused by an inability to reconcile an internal narcissistic identity with an external perceived identity. It should be acknowledged that defence of self-coherence is a goal shared by non-narcissistic behavioural systems, as maintenance of a stable identity is a fundamental human motivation. However, as narcissistic individuals have a fragile sense of self-esteem and therefore greater vulnerability of self-coherence at baseline (Stucke & Sporer, 2002), this behavioural system is afforded higher priority, and becomes governed by distinctly narcissistic motivations.

As a behavioural system with a high priority goal, it naturally has a lower threshold for switching to hyperactivating or deactivating strategies to ensure the system achieves its goal. These strategies produce narcissistic aggression, as drastic but effective means justified by the ends. If these behaviours are associated with positive qualities - for example if the narcissistic individual does not value being perceived as agreeable, whilst valuing the signalling of respect and strength implicit in aggressive actions - then these strategies have the dual benefit of both defending self-coherence from a potential threat and reinforcing self-coherence through signalling positive qualities (Huesmann, 1998), making narcissistic aggression an efficient strategic choice for this behavioural system.

A more straightforward economic account of this behavioural system highlights how narcissistic aggression can play an influential role in maintaining and precipitating the narcissistic behavioural system. Specifically, defence of self-coherence through narcissistic aggression is an active process that requires considerable investment of social, physical, emotional and cognitive resources (e.g. through exerting aggressive behaviour; curating a desirable social context; experiencing negative emotions; analysing ambiguous threats). The consequence of this is that the more an individual successfully maintains the goal of this behavioural system using these strategies, the more the pursuit of this goal is prioritised by virtue of the sunk costs, and the more hyperactivating and deactivating strategies (i.e aggression) become justified by virtue of this increased priority (Shaver et al., 2011). The narcissistic behavioural system may therefore become more influential and justify the use of more extreme strategies. If so, this cyclic self-reinforcing process may play an important role in the development of violent behaviours and clinically significant forms of narcissism.

Finally, narcissistic aggression may also be a means for creating spaces which are advantageous to these implicit economic considerations. That is to say, hyperactivating strategies may become more effective when aggressive and assertive behaviour is normalised and rewarded, and when dissenting voices are removed. Similarly, deactivating strategies may become more effective when dissenting voices

are avoided in order to confirm and protect the hostile attributions underlying narcissistic aggression. In both cases, narcissistic aggression is both responsible for- and a product of the economic considerations involved in the narcissistic behavioural system more generally.

1.3.iv. Subtype differences in narcissistic aggression

Grandiose and vulnerable subtype differences are an essential part of a complete description of narcissism as a trait, and highlight the nuanced nature of narcissistic behaviour and cognition beyond simply exaggerated positive self-beliefs. The influence of these subtype differences with regard to narcissistic aggression is not well understood. However, the theoretical account of narcissistic aggression described here frames subtype differences as emerging in the context of the narcissistic behavioural system, with trait grandiosity or vulnerability being associated with different respective tendencies towards using either hyperactivating or deactivating strategies depending on the situation. For example, a challenging social situation may be associated with an assertive hyperactivating strategy in grandiose narcissistic aggression, whereas in vulnerable narcissistic aggression it may be associated with an avoidant deactivating threat response. Likewise, an insult from a perceived social inferior may be associated with a hostile and dismissive deactivating strategy in grandiose narcissistic aggression, but an explosive and reactive hyperactivating strategy in vulnerable narcissistic aggression. A simpler description may be that grandiose narcissism is associated with using narcissistic aggression to defend self-coherence through creating positive social signalling opportunities (given positive qualities associated with aggression), whereas vulnerable narcissism is associated with using narcissistic aggression in a more pure protective capacity. Grandiose and vulnerable subtypes in narcissistic aggression are therefore both characterised by the pursuit of the goal of defending self-coherence, but reflect different strategic approaches to doing so.

1.4. Summary

Both aggression and narcissism feature rich theoretical descriptions, where an interplay of behavioural, cognitive, affective and motivational forces must be considered in order to develop a satisfactory account of their diverse nature. Despite this, both can be represented in terms of relatively straightforward core mechanisms; aggression as a cycle adjacent to a behavioural system introducing motivational factors, and narcissism as a set of defensive strategies to maintain and protect a fragile sense of self-esteem. When combined, an account of narcissistic aggression is produced wherein narcissism becomes a factor exerting widespread influence across processes of aggression, both through moderating the cycle of aggression and introducing a complex set of motivations. This characterizes both the mechanistic description of how aggression is situated within the narcissistic toolkit of behaviour and cognition, and the practical justification for why narcissistic individuals become aggressive in terms of the economic considerations relevant to narcissistic motivations. Importantly, further nuance is introduced when considering the divergent influence of narcissistic subtype traits on aggressive behaviour, cognition and motivation, which shape how narcissistic aggression varies as a result of individual differences within trait narcissism. The account described here represents an attempt to synthesize insights from across the research literature described above, but this synthesis is novel and therefore represents an original theoretical description of narcissistic aggression. Therefore, in addition to providing a statement of the theoretical approaches to aggression and narcissism featured in subsequent chapters, the above review also provides a formal theoretical framework of narcissistic aggression that will guide the design and interpretation of the empirical component of the thesis.

References

Ackerman, R. A., Witt, E. A., Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., & Kashy, D. A. (2011). What Does the Narcissistic Personality Inventory Really Measure? *Assessment*, *18*(1), 67–87. https://doi.org/10.1177/1073191110382845

Allen, J. J., Anderson, C. A., & Bushman, B. J. (2018). The General Aggression Model. *Current Opinion in Psychology*, *19*, 75–80. https://doi.org/10.1016/j.copsyc.2017.03.034

Anderson, C. A., & Bushman, B. J. (2002). Human Aggression. *Annual Review of Psychology*, *53*(1), 27–51. https://doi.org/10.1146/annurev.psych.53.100901.135231

Archer, J. (2004). Sex Differences in Aggression in Real-World Settings: A Meta-Analytic Review. *Review of General Psychology*, *8*(4), 291–322. https://doi.org/10.1037/1089-2680.8.4.291

Archer, J., & Coyne, S. M. (2005). An Integrated Review of Indirect, Relational, and Social Aggression. *Personality and Social Psychology Review*, *9*(3), 212–230. https://doi.org/10.1207/s15327957pspr0903_2

Ayduk, O., Gyurak, A., & Luerssen, A. (2008). Individual differences in the rejection-aggression link in the hot sauce paradigm: The case of Rejection Sensitivity. *Journal of Experimental Social Psychology*, *44*(3), 775–782. https://doi.org/10.1016/j.jesp.2007.07.004

Back, M., Küfner, A., Dufner, M., Gerlach, T., Rauthmann, J., & Denissen, J. (2013). Narcissistic Admiration and Rivalry: Disentangling the Bright and Dark Sides of Narcissism. *Journal of Personality and Social Psychology*, *105*, 1013–1037. https://doi.org/10.1037/a0034431

Baumeister, R. F., Bushman, B. J., & Campbell, W. K. (2000). Self-Esteem, Narcissism, and Aggression: Does Violence Result from Low Self-Esteem or from Threatened Egotism? *Current Directions in Psychological Science*, *9*(1), 26–29.

Baumeister, R. F., & Vohs, K. D. (2001). Narcissism as addiction to esteem. *Psychological Inquiry*, *12*(4), 206–210.

Bell, P. A., & Baron, R. A. (1976). Aggression and Heat: The Mediating Role of Negative Affect1. *Journal of Applied Social Psychology*, *6*(1), 18–30. https://doi.org/10.1111/j.1559-1816.1976.tb01308.x

Benson, A. J., Jeschke, J., Jordan, C. H., Bruner, M. W., & Arnocky, S. (2019). Will they stay or will they go? Narcissistic admiration and rivalry predict ingroup affiliation and devaluation. *Journal of Personality*, *87*(4), 871–888. https://doi.org/10.1111/jopy.12441

Boccadoro, S., Wagels, L., Henn, A. T., Hüpen, P., Graben, L., Raine, A., & Neuner, I. (2021). Proactive vs. Reactive Aggression Within Two Modified Versions of the Taylor Aggression Paradigm. *Frontiers in Behavioral Neuroscience*, *15*, 749041. https://doi.org/10.3389/fnbeh.2021.749041

Bogart, L. M., Benotsch, E. G., & Pavlovic, J. D. P. (2004). Feeling Superior but Threatened: The Relation of Narcissism to Social Comparison. *Basic and Applied Social Psychology*, *26*(1), 35–44. https://doi.org/10.1207/s15324834basp2601_4

Bond, M. H. (2004). Culture and Aggression—From Context to Coercion. *Personality and Social Psychology Review*, *8*(1), 62–78. https://doi.org/10.1207/s15327957pspr0801_3

Bowlby, J. (1982). *Attachment: Second Edition*. Basic Books.

Brailovskaia, J., Rohmann, E., Bierhoff, H.-W., & Margraf, J. (2020). The anxious addictive narcissist: The relationship between grandiose and vulnerable narcissism, anxiety symptoms and Facebook Addiction. *PLOS ONE*, *15*(11), e0241632. https://doi.org/10.1371/journal.pone.0241632

Brown, R. P., Budzek, K., & Tamborski, M. (2009). On the Meaning and Measure of Narcissism. *Personality and Social Psychology Bulletin*, *35*(7), 951–964. https://doi.org/10.1177/0146167209335461

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology*, *75*(1), 219–229. https://doi.org/10.1037//0022-3514.75.1.219

Campbell, W. K., Rudich, E. A., & Sedikides, C. (2002). Narcissism, Self-Esteem, and the Positivity of Self-Views: Two Portraits of Self-Love. *Personality and*

Social Psychology Bulletin, *28*(3), 358–368. https://doi.org/10.1177/0146167202286007

Chester, D. S., & DeWall, C. N. (2016). The pleasure of revenge: Retaliatory aggression arises from a neural imbalance toward reward. *Social Cognitive and Affective Neuroscience*, *11*(7), 1173–1182. https://doi.org/10.1093/scan/nsv082

Cohen, D., Nisbett, R. E., Bowdle, B. F., & Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: An 'experimental ethnography.' *Journal of Personality and Social Psychology*, *70*(5), 945–960. https://doi.org/10.1037/0022-3514.70.5.945

DeWall, C. N., & Anderson, C. A. (2011). The general aggression model. In *Human aggression and violence: Causes, manifestations, and consequences* (pp. 15–33). American Psychological Association. https://doi.org/10.1037/12346-001

Dickinson, K. A., & Pincus, A. L. (2003). Interpersonal analysis of grandiose and vulnerable narcissism. *Journal of Personality Disorders*, *17*(3), 188–207. https://doi.org/10.1521/pedi.17.3.188.22146

Dodge, K. A. (1991). The structure and function of reactive and proactive aggression. In *The development and treatment of childhood aggression* (pp. 201–218). Lawrence Erlbaum Associates, Inc.

Emmons, R. A. (1984). Factor analysis and construct validity of the Narcissistic Personality Inventory. *Journal of Personality Assessment*, *48*(3), 291–300. https://doi.org/10.1207/s15327752jpa4803_11

Engelhardt, C. R., & Bartholow, B. D. (2013). Effects of Situational Cues on Aggressive Behavior. *Social and Personality Psychology Compass*, *7*(10), 762–774. https://doi.org/10.1111/spc3.12067

Finkel, E. J. (2007). Impelling and Inhibiting Forces in the Perpetration of Intimate Partner Violence. *Review of General Psychology*, *11*(2), 193–207. https://doi.org/10.1037/1089-2680.11.2.193

Fossati, A., Borroni, S., Eisenberg, N., & Maffei, C. (2010). Relations of proactive and reactive dimensions of aggression to overt and covert narcissism in nonclinical adolescents. *Aggressive Behavior*, *36*(1), 21–27. https://doi.org/10.1002/ab.20332

Gebauer, J. E., & Sedikides, C. (2018). Communal Narcissism: Theoretical and Empirical Support. In A. D. Hermann, A. B. Brunell, & J. D. Foster (Eds.), *Handbook of Trait Narcissism: Key Advances, Research Methods, and Controversies* (pp. 69–77). Springer International Publishing. https://doi.org/10.1007/978-3-319-92171-6_7

Geen, R. G., & Powers, P. C. (1971). Shock and Noise as Instigating Stimuli in Human Aggression. *Psychological Reports*, *28*(3), 983–985. https://doi.org/10.2466/pr0.1971.28.3.983

Glover, N., Miller, J. D., Lynam, D. R., Crego, C., & Widiger, T. A. (2012). The five-factor narcissism inventory: A five-factor measure of narcissistic personality traits. *Journal of Personality Assessment*, *94*(5), 500–512. https://doi.org/10.1080/00223891.2012.670680

Grapsas, S., Brummelman, E., Back, M. D., & Denissen, J. J. A. (2020a). The "Why" and "How" of Narcissism: A Process Model of Narcissistic Status Pursuit. *Perspectives on Psychological Science*, *15*(1), 150–172. https://doi.org/10.1177/1745691619873350

Grapsas, S., Brummelman, E., Back, M. D., & Denissen, J. J. A. (2020b). The "Why" and "How" of Narcissism: A Process Model of Narcissistic Status Pursuit. *Perspectives on Psychological Science, 15*(1), 150–172. https://doi.org/10.1177/1745691619873350

Hendin, H. M., & Cheek, J. M. (1997). Assessing hypersensitive narcissism: A reexamination of Murray's Narcism Scale. *Journal of Research in Personality*, *31*(4), 588–599. https://doi.org/10.1006/jrpe.1997.2204

Hogg, M. A., & Cooper, J. (2007). *The SAGE Handbook of Social Psychology: Concise Student Edition*. SAGE.

Huesmann, L. R. (1998). 4—The Role of Social Information Processing and Cognitive Schema in the Acquisition and Maintenance of Habitual Aggressive Behavior. In R. G. Geen & E. Donnerstein (Eds.), *Human Aggression* (pp. 73–109). Academic Press. https://doi.org/10.1016/B978-012278805-5/50005-5

Huesmann, L. R., & Kirwil, L. (2007). Why Observing Violence Increases the Risk of Violent Behavior By the Observer. In A. T. Vazsonyi, D. J. Flannery, & I. D.

Waldman (Eds.), *The Cambridge Handbook of Violent Behavior and Aggression* (pp. 545–570). Cambridge University Press. https://doi.org/10.1017/CBO9780511816840.029

Huesmann, L. R., Moise-Titus, J., Podolski, C.-L., & Eron, L. D. (2003). Longitudinal relations between children's exposure to TV violence and their aggressive and violent behavior in young adulthood: 1977-1992. *Developmental Psychology*, *39*(2), 201–221. https://doi.org/10.1037/0012-1649.39.2.201

Hurlbert, D. F., Apt, C., Gasar, S., Wilson, N. E., & Murphy, Y. (1994). Sexual narcissism: A validation study. *Journal of Sex & Marital Therapy*, *20*(1), 24–34. https://doi.org/10.1080/00926239408403414

Jauk, E., Weigle, E., Lehmann, K., Benedek, M., & Neubauer, A. C. (2017). The Relationship between Grandiose and Vulnerable (Hypersensitive) Narcissism. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01600

Kampe, L., Bohn, J., Remmers, C., & Hörz-Sagstetter, S. (2021). It's Not That Great Anymore: The Central Role of Defense Mechanisms in Grandiose and Vulnerable Narcissism. *Frontiers in Psychiatry*, *12*, 661948. https://doi.org/10.3389/fpsyt.2021.661948

Kealy, D., & Rasmussen, B. (2012). Veiled and Vulnerable: The Other Side of Grandiose Narcissism. *Clinical Social Work Journal*, *40*(3), 356–365. https://doi.org/10.1007/s10615-011-0370-1

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review*, *110*(2), 265–284. https://doi.org/10.1037/0033-295X.110.2.265

Kim, H., & Barry, C. T. (2021). The moderating effect of intolerance of uncertainty on the relation between narcissism and aggression. *Current Psychology*. https://doi.org/10.1007/s12144-021-01542-9

Levy, K. N., Ellison, W. D., & Reynoso, J. S. (2011). A historical review of narcissism and narcissistic personality. In *The handbook of narcissism and narcissistic personality disorder: Theoretical approaches, empirical findings, and treatments* (pp. 3–13). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118093108

Lickley, R. A., & Sebastian, C. L. (2018). The neural basis of reactive aggression and its development in adolescence. *Psychology, Crime & Law*, *24*(3), 313–333. https://doi.org/10.1080/1068316X.2017.1420187

Lieberman, J. D., Solomon, S., Greenberg, J., & McGregor, H. A. (1999). A hot new way to measure aggression: Hot sauce allocation. *Aggressive Behavior*, *25*(5), 331–348. https://doi.org/10.1002/(SICI)1098-2337(1999)25:5<331::AID-AB2>3.0.CO;2-1

Liu, J. (2004). Concept analysis: Aggression. *Issues in Mental Health Nursing*, *25*(7), 693–714. https://doi.org/10.1080/01612840490486755

McCarthy, R. J., & Elson, M. (2018). A Conceptual Review of Lab-Based Aggression Paradigms. *Collabra: Psychology*, *4*(1), 4. https://doi.org/10.1525/collabra.104

Miller, J. D., Hoffman, B. J., Gaughan, E. T., Gentile, B., Maples, J., & Keith Campbell, W. (2011). Grandiose and vulnerable narcissism: A nomological network analysis. *Journal of Personality*, *79*(5), 1013–1042. https://doi.org/10.1111/j.1467-6494.2010.00711.x

Miller, J. D., & Maples, J. (2011). Trait personality models of narcissistic personality disorder, grandiose narcissism, and vulnerable narcissism. In *The handbook of narcissism and narcissistic personality disorder: Theoretical approaches, empirical findings, and treatments* (pp. 71–88). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118093108

Miller, N., Pedersen, W. C., Earleywine, M., & Pollock, V. E. (2003). A Theoretical Model of Triggered Displaced Aggression. *Personality and Social Psychology Review*, *7*(1), 75–97. https://doi.org/10.1207/S15327957PSPR0701_5

Morf, C. C., & Rhodewalt, F. (2001). Unraveling the Paradoxes of Narcissism: A Dynamic Self-Regulatory Processing Model. *Psychological Inquiry*, *12*(4), 177–196. https://doi.org/10.1207/S15327965PLI1204_1

Okada, R. (2010). The relationship between vulnerable narcissism and aggression in Japanese undergraduate students. *Personality and Individual Differences*, *49*(2), 113–118. https://doi.org/10.1016/j.paid.2010.03.017

Raskin, R. N., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports, 45*(2), 590–590. https://doi.org/10.2466/pr0.1979.45.2.590

Raskin, R., Novacek, J., & Hogan, R. (1991). Narcissistic self-esteem management. *Journal of Personality and Social Psychology*, *60*(6), 911–918. https://doi.org/10.1037/0022-3514.60.6.911

Reidy, D. E., Foster, J. D., & Zeichner, A. (2010). Narcissism and unprovoked aggression. *Aggressive Behavior*, *36*(6), 414–422. https://doi.org/10.1002/ab.20356

Roberts, R., Callow, N., Hardy, L., Woodman, T., & Thomas, L. (2010). Interactive Effects of Different Visual Imagery Perspectives and Narcissism on Motor Performance. *Journal of Sport & Exercise Psychology*, *32*, 499–517. https://doi.org/10.1123/jsep.32.4.499

Shaver, P. R., Segev, M., & Mikulincer, M. (2011). *A behavioral systems perspective on power and aggression.* 71. https://doi.org/10.1037/12346-004

Singh, T., Mohanty, R., Lalrhiatpuia, L., Haobam, M., & Saini, M. (2014). Aggression and Violent Behaviour: A Critical Review. *IOSR Journal of Pharmacy and Biological Sciences*, *9*, 10–13. https://doi.org/10.9790/3008-09551013

Slotter, E. B., & Finkel, E. J. (2011). *I^3 theory: Instigating, impelling, and inhibiting factors in aggression.* 35. https://doi.org/10.1037/12346-002

Stucke, T. S., & Sporer, S. L. (2002). When a grandiose self-image is threatened: Narcissism and self-concept clarity as predictors of negative emotions and aggression following ego-threat. *Journal of Personality*, *70*(4), 509–532. https://doi.org/10.1111/1467-6494.05015

Szymczak, P., Sawicki, A., & Jaworski, M. (2020). How narcissists see the social world? Trust, cynicism, and trifurcated model of narcissism. *Current Psychology*. https://doi.org/10.1007/s12144-020-01215-z

Vater, A., Moritz, S., & Roepke, S. (2018). Does a narcissism epidemic exist in modern western societies? Comparing narcissism and self-esteem in East and West Germany. *PLoS ONE*, *13*(1). https://doi.org/10.1371/journal.pone.0188287

Vitaro, F., & Brendgen, M. (2005). Proactive and Reactive Aggression: A Developmental Perspective. In *Developmental origins of aggression* (pp. 178–201). The Guilford Press.

Wallace, H. M., & Baumeister, R. F. (2002). The performance of narcissists rises and falls with perceived opportunity for glory. *Journal of Personality and Social Psychology*, *82*(5), 819–834. https://doi.org/10.1037/0022-3514.82.5.819

Watts, A. L., Lilienfeld, S. O., Smith, S. F., Miller, J. D., Campbell, W. K., Waldman, I. D., Rubenzer, S. J., & Faschingbauer, T. J. (2013). The Double-Edged Sword of Grandiose Narcissism: Implications for Successful and Unsuccessful Leadership Among U.S. Presidents. *Psychological Science*, *24*(12), 2379–2389. https://doi.org/10.1177/0956797613491970

Wilkowski, B. M., & Robinson, M. D. (2008). The cognitive basis of trait anger and reactive aggression: An integrative analysis. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *12*(1), 3–21. https://doi.org/10.1177/1088868307309874

Wink, P. (1991). Two faces of narcissism. *Journal of Personality and Social Psychology*, *61*(4), 590–597. https://doi.org/10.1037//0022-3514.61.4.590

Winnicott, D. W. (1984). Ego Distortion in Terms of True and False Self (1960). In *The Maturational Processes and the Facilitating Environment*. Routledge.

Yakeley, J. (2018). Current understanding of narcissism and narcissistic personality disorder. *BJPsych Advances*, *24*(5), 305–315. https://doi.org/10.1192/bja.2018.20

Zajenkowski, M., Maciantowicz, O., Szymaniak, K., & Urban, P. (2018). Vulnerable and Grandiose Narcissism Are Differentially Associated With Ability and Trait Emotional Intelligence. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01606

Zeigler-Hill, V., McCabe, G. A., Vrabel, J. K., Raby, C. M., & Cronin, S. (2018). The Narcissistic Pursuit of Status. In A. D. Hermann, A. B. Brunell, & J. D. Foster (Eds.), *Handbook of Trait Narcissism: Key Advances, Research Methods, and Controversies* (pp. 299–306). Springer International Publishing. https://doi.org/10.1007/978-3-319-92171-6_32

2. Social pain: narcissistic sensitivity to rejection

2.1. Introduction

The existing literature documenting narcissism and narcissistic aggression provides a sufficiently clear overall theoretical account, allowing assumptions to be made regarding the beliefs and motivations underlying narcissistic cognition and behavior and predictions to be made regarding where and how this cognition and behaviour should manifest. However, the literature is less clear as a resource when it comes to making practical decisions regarding an appropriate methodology for studying narcissistic aggression. There have been issues with the use of psychometric tools for studying narcissism. For example, the somewhat inconsistent use of the NPI to represent either a general trait narcissism or more specifically grandiose narcissism (Chrétien et al., 2018; Raskin & Hall, 1979), or the concerns regarding the construct validity of the HSNS insofar as it seeks to measure more than just sensitivity (Hendin & Cheek, 1997). Furthermore, until relatively recently there has been only infrequent acknowledgement of narcissistic subtypes, limiting the extent to which the psychometric literature can be considered compatible with the contemporary conception of narcissism. Similarly, the variety of different experimental operationalisations of aggression used in the literature limits the extent to which results can be considered compatible across the literature (McCarthy & Elson, 2018). For example, it is uncertain whether one can assume that the type of aggression elicited by delivering an irritating stimulus is the same as that elicited by a personal insult.

In a practical sense, narcissistic aggression is a psychological concept that does not exist independently of the methods used to measure it, and so it is best understood within the context of a specific set of measures. As a result, whilst this lack of clarity in the literature does not present a limitation to broad theoretical judgements concerning narcissism and aggression, it does present a limitation when attempting to

make more precise experimental predictions, such as those concerning the particular differences between narcissistic subtypes and the particular characteristics of narcissistic aggression in a given context. For example, whilst there are experimental studies that examine grandiose narcissistic aggression (Bushman et al., 2009; Reidy et al., 2010; Seah & Ang, 2008) and vulnerable narcissistic aggression (Okada, 2010) independently, it is particularly valuable to study both grandiose and vulnerable narcissistic aggression using the same operationalisation of aggression, examples of which are rare in the existing literature.

In light of this, the present study aimed to select and test a method that examined several aspects of narcissistic aggression in the context of the same operationalisation of aggression, in order to provide data that is less biased by the type of methodological inconsistency present in existing literature, and describe a form of narcissistic aggression particular to its context (with acknowledgement of the usefulness and limitations of this description of narcissistic aggression). More generally, the present study aimed to provide insight into how narcissistic aggression manifests in the environment of a laboratory, in order to facilitate future experiments in this series that may use different methods but be similarly laboratory based.

The first part of designing this method is to select a suitable measure for narcissism. To address issues of construct validity with respect to the psychometric tools described above, the FFNI (Glover et al., 2012) is considered the most appropriate measure (see **Chapter 1** for a discussion of the different psychometric measures of narcissism).

The second and more challenging part is selecting an appropriate operationalisation for aggression. Unlike the identified methodological issues concerning narcissism, which primarily involve construct validity and are relatively straightforward to resolve, the methodological issues concerning aggression are rooted in the diversity of potential ways in which one can operationalise aggression, and the difficulty of making a non-arbitrary choice of method. However, the theoretical framework of

narcissistic aggression developed earlier in the thesis (see **Chapter 1**) can assist with this choice. It is assumed that at the root of narcissistic aggression (and indeed narcissism more generally) is a fragile sense of self esteem (Zeigler-Hill & Jordan, 2011) that makes a narcissistic individual vulnerable to stressful states of decoherent identity, where external evidence threatens internal narcissistic self-beliefs. Aggressive behavioral and cognitive strategies present an effective means of defending this vulnerability, and may provide an opportunity to assert traits that are perceived as positive and more coherent with a narcissistic identity. As a result, a method for eliciting aggression based on a threat to self-esteem is an appropriate choice. Furthermore, as narcissistic aggression is archetypally social (insofar as occurring mostly exclusively in social groups is a defining feature) then it is appropriate that this threat to self-esteem involves a social insult (i.e. a direct or indirect negative evaluation from another individual).

A widely used experimental method for simulating a social insult that implicitly threatens self-esteem is the Cyberball paradigm (Williams & Jarvis, 2006; for examples see Hartgerink et al., 2015). Cyberball is a simple behavioral task in which the participant plays a virtual cooperative ball throwing video game with two or more other participants. The other participants are simulated confederates, and during the task begin to selectively exclude the real participant from the game (i.e. by avoiding passing the ball to the real participant). The overt and protracted nature of the exclusion provides reasonable grounds for any participant to interpret it as a social insult, and so narcissistic participants in particular should be likely to interpret this exclusion as an act of deliberate rejection with the intent to signal disapproval of the participant. This paradigm is highly customisable, and implementing an operationalisation for aggression into the task is straightforward, allowing the task to act as a means of both inducing and measuring aggression. As narcissistic aggression features both hyperactivating (i.e. assertive and confrontational) and deactivating (i.e. avoidant and hostile) strategies, this modified Cyberball paradigm should include opportunities for both strategies after the social insult takes place. In the present study, a hyperactivating opportunity is presented in the form of a follow-up task in

which the cooperative portion of the game is followed by a competitive task where the participant can compete with the virtual confederates and earn the chance to deliver an uncomfortable noise blast if they win. This opportunity is also given to participants prior to the cooperative task to selectively examine any tendency for assertive aggressive behaviour independent of provocation. A deactivating opportunity is presented in the form of an opportunity to prematurely end the experiment after the cooperative task but before the competitive task, resulting in the experiment also ending prematurely for the virtual confederates but without any other penalty. An operationalisation of aggression featuring the ability to measure use of both these strategies is particularly relevant in the context of grandiose and vulnerable subtype differences in narcissistic aggression. Finally, in order to add redundancy to this behavioral operationalisation of aggression, relevant psychometric measures should also be included. In the present study, these are the Situational Triggers of Aggressive Responses Scale (STAR; Lawrence, 2006) and the Behaviour Approach/Inhibition System Scale (BIS/BAS; Carver & White, 1994), which respectively measure trait threat sensitivity and disposition towards approach (i.e. hyperactivating) and avoidance (i.e. deactivating) behaviours, as well as the UWIST Mood Adjective Checklist (UMACL; Matthews et al., 1990) to examine state mood that may be relevant to aggression.

This is therefore a theoretically appropriate operationalisation of narcissistic aggression. The present study aims to explore the method described above to investigate its viability and whether the type of operationalisation featured in this method might be useful for future experiments in this series. This includes both the behavioural task featured, as well as the viability of the FFNI and its relationships with aggressive behaviour and relevant traits. Finally, and no less importantly, the present study aims to further the understanding of narcissistic aggression and the significance of subtype differences, when elicited by mild provocation in a laboratory setting.

Hypotheses

The present study is partly exploratory, taking a relatively open approach to investigating how narcissism as defined by the FFNI relates to aggression elicited by this Cyberball paradigm, how the FFNI more generally relates to forms of aggression occurring in laboratory settings featuring mild provocation via social insult, and indeed whether this is an effective operationalisation of aggression in practice. Additionally, general state mood is assessed (using the UMACL questionnaire) as an exploratory measure and a potential experimental confound influencing threat-related behaviour.

However, the present study (and subsequent studies) also investigates several more defined hypotheses based on the assumptions of the theoretical framework of narcissistic aggression outlined in the first chapter. Specifically, higher trait grandiose narcissism should be associated with greater likelihood to deliver a noise blast in response to the provocation featured in the Cyberball task (i.e. the follow-up noise blast opportunity), given the theorised grandiose narcissistic disposition towards assertive hyperactivating behavioural strategies. Likewise, higher trait grandiose narcissism should be associated with a greater desire to deliver a noise blast. Conversely, higher trait vulnerable narcissism should be associated with a greater desire to withdraw from the experiment following the provocation featured in the Cyberball task, given the theorised vulnerable narcissistic disposition towards avoidant deactivating behavioural strategies. This should also be reflected in a lower desire to receive a noise blast.

Similarly, higher trait grandiose narcissism is hypothesized to be associated with higher trait behavioural approach, whereas higher trait vulnerable narcissism should be associated with higher trait behavioural inhibition (as measured by BIS/BAS). Both higher trait grandiose and vulnerable narcissism are hypothesized to be associated with higher trait sensitivity to provocations (as measured by STAR), given

the theorised shared sensitivity to threats that exists irrespective of different behavioural tendencies.

Finally, if the operationalisations of threat-related behaviour and related cognition featured in the present study (i.e. measures relating to participating in the competitive task featuring a noise blast) are functional, it is expected that they should be associated with appropriate trait measures. Specifically, higher trait sensitivity to provocations should be associated with a higher likelihood to deliver a noise blast following the provocation event, and a greater willingness to deliver a noise blast. Similarly, trait behavioural approach and behavioural inhibition should be associated with a respective increase or decrease in willingness to continue to participate after the provocation.

2.2. Method

Participants

Participants were 31 students (10 male, 21 female) aged between 19 and 37 years (*M* = 22.5) attending the University of Nottingham, who had volunteered to participate in experiments within the School of Psychology. This sample size is consistent with other exploratory studies based on the Cyberball paradigm (Bernstein & Claypool, 2012a, 2012b; Jamieson et al., 2010; Sacco et al., 2011). Participants were told that they would be taking part in a study to examine how individuals behave socially when playing multiplayer computer games, and would be required to play several games with two other anonymous participants. Participants gave their informed consent and were provided with an inconvenience allowance for their participation. The experiment was approved by the School of Psychology ethics committee (Ref: S995R).

Procedure

Participants arrived in a group size between one and six, and were seated in separate private cubicles within the same room, each cubicle containing desk space and a computer. As the main task involved each participant believing that they were playing against two other players (in a triad), a group size of six was optimum to ensure participants could not be certain who they were playing against (i.e. who was in their triad). Participants were hidden from one another for this same reason, so that they could not guess their partners by studying the reactions of others. When it was not possible to achieve a group size of six on a given day, the researcher would explain to participants that they were but one of two groups participating in the experiment that day, and that they would be playing remotely with participants from the other group (which was suggested to be of sufficient size that any vacancies in the participant's group could be compensated).

Participants were required to complete the Five-Factor Narcissism Inventory Short Form (FFNI-SF; Sherman et al., 2015), the Situational Triggers of Aggressive Responses Scale (STAR) Scale, and the Behaviour Approach/Inhibition System Scale (BAS/BIS). They then participated in three behavioural tasks.

First, participants played a twenty-trial competitive reaction task game, ostensibly with an anonymized triad of players, where they were required to hit the spacebar when they saw a stimulus appear on screen, with the player who achieved the fastest average reaction time being declared the winner. Participants were informed prior to starting the game that each player could elect to deliver a noise blast to the two losing players in the event that they won, and were required to make this decision prior to playing each trial. If a participant chose to deliver a noise blast in the event they won the game, they were also able to choose the intensity and duration of the noise blast. As each participant was actually playing alone, their decision whether or not to deliver the noise blast had no real consequence, although the participant was not aware of this. In this initial task, the participant was always pre-programmed to be the 'winner'.

Second, participants played the Cyberball game with an anonymized triad of players that they were informed were different from the triad they had played with in the previous task. To ensure engagement in the task, participants were informed prior to starting the game that they would need to make at least 100 passes as a group, a number easily achievable in the time available. During the first half of the game, the two computer controlled players in the triad behave normally and pass to the participant as frequently as they pass to one another. However, in the latter half of the game the two computer controlled players begin passing to each other more frequently, and shortly after begin ignoring the participant altogether. During the exclusion period, the other 'players' make only two pre-programmed passes to the participant. These two passes were intentionally added to ensure that participants did not conclude that their exclusion was non-intentional (e.g. a computer glitch) or the product of task disengagement (e.g. simply other 'players' simply pressing the same button repeatedly). When this game was complete, participants were required to complete the UWIST Mood Adjective Checklist (UMACL).

Finally, participants played a second version of the initial competitive reaction task, and were informed that they would be playing with the *same* triad of players that participated with them in the ball throwing game. The procedure for this task was the same as the initial task, but prior to starting participants were asked how happy they were to participate in this second instance of the reaction time game, how happy they were that the losing players should receive a noise blast, and how happy they were with having the ability to deliver a noise blast should they win. As in the first reaction time task, each participant was pre-programmed to be the 'winner' regardless of their performance. The experiment ended with each participant viewing a thirty second positive mood primer video to mitigate any upset negative affect they may have experienced as a result of the experiment.

Behavioral tasks

The behavioural tasks were presented as a series of virtual games. In all cases, participants were led to believe they were playing in a triad of real participants, but in reality each participant was playing on their own with fake bogus (reaction time games) or computer simulated (ball throwing game) partners.

Reaction time games. Both the first and second reaction time tasks consisted of a blue screen, upon which a small yellow circle appeared (the reaction stimulus). Upon pressing the spacebar, the yellow circle would disappear for an interval between 0.5 and 2 seconds long, after which it would appear again. This repeated for 20 trials. Whilst reaction times were recorded for each participant, there was no actual benchmark for performance and no penalty for reacting prematurely (i.e. during an interval).

Ball throwing game. The game was a modified version of the Cyberball task (Williams & Jarvis, 2006) in which each player takes turns passing a virtual ball between each their own and other 's players' respective virtual avatars. The game is presented as a blue screen with the names 'Player A' and 'Player B' visible in the top corners of the screen, and the name 'You' visible in the bottom middle. The ball is represented by a small image of a football. The participant (i.e. 'You') starts with the ball positioned next to their name, and upon pressing either 'A' or 'B' on the keyboard, the ball is relocated next to the name of either Player A or B. This then repeats depending on the decision pre-programmed protocol made by the computer for Player A or B, and so on. Participants are kept informed of their own actions and the actions of the other player with text prompts (e.g. "Player A passed to you!"). This repeated for 115 trials or 'passes'. For the first 70 trials, the computer was programmed to have an equal chance of passing to either the participant or another computer-controlled player. Between trials 70 and 80, the computer began partially excluding the participant, with a 66% chance of passing to the each of the other computer-controlled players. Between trials 80 and 115, the computer excluded the participant, passing solely to the computer-controlled player, with the exception of two pre-programmed passes to the participant at trial 90 and trial 110.

All of the behavioural tasks used were programmed using the PsychoPy Coder (V. 1.850) programming framework for Python.

Psychological questionnaire measures

Narcissism. Both grandiose and vulnerable narcissism were assessed using the 60-item Five-Factor Narcissism Inventory Short Form (FFNI-SF; Sherman et al., 2015). Participants responded to each statement on a 5 point Likert scale ranging from 1 (disagree strongly) to 5 (agree strongly). The questionnaire consists of two primary sub-scales containing 44 items assessing grandiose narcissism (e.g. "Others say I brag too much, but everything I say is true") and 16 items assessing vulnerable narcissism (e.g. "I often feel as if I need compliments from others to be sure of myself") respectively.

Threat sensitivity. The Situational Triggers of Aggressive Responses Scale (STAR; Lawrence, 2006) was used to assess how sensitive participants were to aggressive triggers (specifically: provocations and frustrations). Participants responded to 22 statements on a 5 point Likert scale ranging from 1 (very inaccurate) to 5 (very accurate). The questionnaire consists of two sub-scales assessing sensitivity to provocations (12 items e.g. "Someone insults me") and frustrations (e.g. "I hear a noise I cannot control").

Behavioural approach and inhibition. The Behaviour Approach/Inhibition System Scale (BAS/BIS; Carver & White, 1994) was used to assess how participants regulated approach and inhibition behaviours. The questionnaire consists of 24 items assessing aspects of behavioural approach (e.g. "I often act on the spur of the moment") and inhibition (e.g. "I worry about making mistakes"). Participants responded to each item on a 4 point Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree).

Mood. The UWIST Mood Adjective Checklist (UMACL; Matthews et al., 1990) was used to assess participant state mood. The questionnaire consists of 48 single-word items assessing different aspects of mood, specifically hedonic tone, anger, tense arousal and energetic arousal. Three additional items were added to assess mood states of particular interest in the study, specifically "isolated", "humiliated" and "social". Participants rated how much each item applied to them using to a 4 point Likert scale ranging from 1 (definitely not) to 4 (definitely).

Willingness to continue scale. An original measure was designed for this study to assess participants' attitudes towards taking part in the second reaction time task. The measure consisted of three questions (i) "How happy are you to participate in this second instance of the reaction time game?" (ii), "How happy are you with the possibility for a player to deliver a noise blast to the losing players in the event that they win? (iii) "How happy are you with the possibility that losing players could receive a noise blast in the event that they lose?". Participants gave their rating on a 10-point Likert scale ranging from 1 (Very Unhappy) to 10 (Very Happy).

Data Analysis

Data from three participants was corrupted due to a programming error affecting data collected from the ball-throwing game onwards, meaning that they could only be included in analyses featuring data collected up to and including the first instance of the reaction time task. As a result, any analyses after this point have a reduced sample size (N = 28).

An independent-samples Mann-Whitney U Test was conducted to test whether individuals who chose to deliver a noise blast in the first competitive reaction time task, and individuals who chose to do so in the second competitive reaction time task, exhibited different ratings of trait grandiose narcissism and vulnerable narcissism compared to those who did not. Multiple bivariate Pearson's correlation analyses were conducted to investigate the relationship between grandiose narcissism and vulnerable narcissism and the other variables of interest measured in the present study

(i.e. ratings on the STAR, BAS/BIS and UMACL scales), and with the experimental variables (i.e. chosen intensity of the noise blasts pre and post social insult, at both the first and second instance, the difference between pre and post social insult blast intensities, and willingness to participate in the reaction-time task recorded prior to the second competitive reaction time task).

2.3. Results

Difference in GN/VN between those who chose to deliver a noise blast and those who did not

Mann-Whitney U Tests indicated that ratings of grandiose narcissism scores were significantly higher among individuals who chose to deliver a noise blast prior to the first instance of the competitive reaction time task (*Median (Mdn)* = 114, *N* = 11) than individuals who did not (*Mdn* = 104, *N* = 20; *U(31)* = 59, *p* = .035, Pearson's *r (r)* = .37). No significant difference in ratings of vulnerable narcissism was observed between the two groups (*U(31)* = 81, *p* = .230, *r* = .21). No significant difference in ratings of grandiose narcissism was observed between individuals who chose to deliver a noise blast prior to the second instance of the competitive reaction time task (*N* = 8) and individuals who did not (*N* = 20; *U(28)* = 49.5, *p* = .120, *r* = .29), nor was any significant difference in ratings of vulnerable narcissism (*U(28)* = 56.5, *p* = .230, *r* = .22).

The mean, standard deviation, minimum and maximum values, and Cronbach's alpha statistics for each measure are listed in **Table 1**. Correlations between ratings of grandiose narcissism, vulnerable narcissism, and ratings on all sub-scales of the variables of interest and administered questionnaires are listed in **Table 2** and **Table 3**, whereas correlations between grandiose narcissism and vulnerable narcissism, and experimental variables are listed in **Table 4**.

Table 1. Descriptive statistics for measures used.

	Mean	SD	Minimum	Maximum	α
FFNI Grandiose narcissism	108.45	17.26	57	141	.866
FFNI Vulnerable narcissism	51.13	8.87	30	66	.776
STAR Provocations	38.55	6.61	24	52	.676
STAR Frustrations	27.20	7.45	14	45	.807
BAS/BIS Drive	10.42	1.71	7	14	.633
BAS/BIS Fun Seeking	11.87	2.20	7	15	.674
BAS/BIS Reward Responsiveness	17.71	1.92	14	20	.950
BAS/BIS Inhibition	20.77	1.67	17	25	.872
UMACL Happy	17.71	10	24	3.53	.891
UMACL Sad	10.71	7	20	3.46	.856
UMACL Anger	7.96	5	15	3.06	.854
UMACL Aroused	11.54	8	21	3.61	.800
UMACL Calm	27.32	19	36	4.71	.765
UMACL Active	19.75	12	31	4.48	.715
UMACL Inactive	17.75	9	25	4.29	.478
Willingness to continue	7.42	4	10	1.73	*
Willingness to deliver noise blast	3.84	1	9	2.22	*
Willingness to receive noise blast	4.10	1	10	2.52	*

** Scale consisted of a single item*

Table 2. Zero order correlations between grandiose/vulnerable narcissism and variables of interest, including all trait sub-scales (STAR Scale and BAS/BIS).

	FFNI-GN	**FFNI-VN**	**STAR Provocations**	**STAR Frustrations**	**BAS/BIS Drive**	**BAS/BIS Fun Seeking**	**BAS/BIS Reward**	**BAS/BIS Reponsiveness**	**BAS/BIS Inhibition**
FFNI Grandiose narcissism	-	.508**	.482**	.287	.218	.402*	.181	0.69	
FFNI Vulnerable narcissism		-	.479**	.260	.201	.016	.161	.401*	

** p<0.05, ** p<0.01. GN = Grandiose Narcissism, VN = Vulnerable Narcissism*

Table 3. Zero order correlations between grandiose/vulnerable narcissism and variables of interest, including all sub-scalesmood post experimental tasks (UMACL).

	UMACL Happy	**UMACL Sad**	**UMACL Anger**	**UMACL Aroused**	**UMACL Calm**	**UMACL Active**	**UMACL Inactive**
FFNI Grandiose narcissism	.036	.094	.326	.158	.051	.057	.222
FFNI Vulnerable narcissism	.161	.240	.245	.030	.019	.068	.026

For all analyses involving UMACL, N = 28 due to data corruption (see above).

** p<0.05, ** p<0.01. GN = Grandiose Narcissism, VN = Vulnerable Narcissism*

Relationship between GN/VN and control variables

The correlational analyses indicated a strongly significant positive correlation between grandiose and vulnerable narcissism, and the sensitivity to provocations sub-scale of the STAR scale, as well as a significant positive correlation between grandiose narcissism and the fun seeking sub-scale of BAS/BIS. Furthermore, the analyses indicated a strong positive correlation between vulnerable narcissism and sensitivity to provocations, and a significant positive correlation between vulnerable narcissism and with the behavioural inhibition subscale of BAS/BIS. No significant correlations were observed between any other variables of interest.

Table 4. Zero order correlations between grandiose/vulnerable narcissism and experimental variables (noise blast intensities and willingness to participate in the second reaction time task).

	Noise Blast Intensity (1st)	**Noise Blast Intensity (2nd)**	**Noise Blast Intensity (1st - 2nd)**	**Willingness to continue**	**Willingness to give noise blast**	**Willingness to receive noise blast**
FFNI Grandiose narcissism	.192	.188	.115	.045	.442*	.381*
FFNI Vulnerable narcissism	.006	-.139	-.184	.316	.398*	.327

** p<0.05. For all analyses involving the willingness to continue scale, N = 28 due to data corruption (see above). Only participants who chose to deliver a noise blast (N = 11) provided ratings of intensity.*

Table 5. Zero order correlations between variables of interest and trait sensitivity to aggressive triggers, behavioural approach and behavioural inhibition and experimental variables.

	Noise Blast Intensity (1st)	**Noise Blast Intensity (2nd)**	**Noise Blast Intensity (1st - 2nd)**	**Willingness to continue**	**Willingness to give noise blast**	**Willingness to receive noise blast**
STAR Provocations	-.005	-.283	-.556	.083	.452*	.551**
STAR Frustrations	-.058	.025	.168	.100	.244	.260
BAS/BIS Drive	.005	.370	.469	.113	-.011	.008
BAS/BIS Fun Seeking	.438	.402	.120	-.045	-.099	-.174
BAS/BIS Reward	.132	.166	.043	-.079	-.079	-.128
BAS/BIS Inhibition	-.344	-.588	-.441	.372	.056	.005

** p<0.05, ** p<0.01. For all analyses involving the willingness to continue scale, N = 28 due to data corruption (see above). Only participants who chose to deliver a noise blast (N = 11) provided ratings of intensity.*

Relationship between GN/VN, willingness to continue, and noise blast intensity

Correlational analyses between grandiose and vulnerable narcissism and the experimental variables indicated a significant positive correlation between grandiose

narcissism and willingness to deliver a noise blast to losing players and willingness to receive a noise blast as a losing player. Furthermore, analyses indicated a significant positive correlation between vulnerable narcissism and willingness to deliver a noise blast to losing players. No significant correlations were observed between grandiose and vulnerable narcissism and any other experimental variables.

Relationship between control variables and decision to deliver noise blast, willingness to continue and noise blast intensity

Correlations between variables of interest (i.e. those other than grandiose and vulnerable narcissism) and experimental variables are listed in **Table 5.** Correlational analyses between variables of interest and experimental variables indicated a significant positive correlation between the sensitivity to provocations sub-scale of the STAR scale and willingness to deliver a noise blast to losing players, and a strong significant positive correlation with willingness to receive a noise blast as a losing player. No significant correlations were observed.

Mann-Whitney U Tests indicated that ratings scores on the sensitivity to provocations sub-scale of the STAR scale were significantly higher among individuals who chose to deliver a noise blast prior to the first instance of the competitive reaction time task (Mdn = 44, N = 11) than individuals who did not (Mdn = 35, N = 20; $U(31)$ = 44, p = .006, r = .49), and those who chose to deliver a noise blast prior to the second instance of the competitive reaction time task (Mdn = 44.5, N = 8) and individuals who did not (Mdn = 36.5, N = 20; $U(28)$ = 37.5, p = .030, r = .40). No other significant differences were observed between either group, at either instance, for any other variables of interest.

2.4. Discussion

Several results were observed coherent with the hypotheses of the present study. Grandiose narcissism was associated with a greater desire to deliver a noise blast, as well as a greater likelihood to deliver a noise blast (although this association was

independent of provocation, see below for further discussion). Grandiose narcissism was also associated with a trait measure of behavioural approach, as well as trait sensitivity to provocations. Vulnerable narcissism was associated with trait behavioural inhibition, as well as trait sensitivity to provocations. Finally, trait sensitivity to provocations was associated both with a greater desire to deliver a noise blast, and a greater likelihood to deliver a noise blast at both the first and second instance. However, several unexpected or absent effects were also observed.

The findings of the present study therefore provide insight into the nature of narcissistic aggression, presenting several useful methodological considerations and highlighting several interesting grandiose and vulnerable subtype differences in narcissistic aggression outlined below.

Differences between the aggressive behaviour of grandiose and vulnerable narcissists

Individuals who chose to deliver a noise blast prior to the beginning of the first reaction time game exhibited higher ratings of grandiose narcissism than those who did not. Given that this occurred prior to the social exclusion event in the cooperative ball-throwing game, the decision to administer an unpleasant stimulus here constitutes unprovoked aggressive behaviour. This association between unprovoked aggression and grandiose narcissism has been suggested elsewhere; grandiose narcissism is associated with more readily delivering an electric shock irrespective of provocation (Reidy et al., 2010) and criminal offenders with higher grandiose narcissism exhibit higher ratings of psychopathic traits relating to life history of aggression, particularly unprovoked assaults (Schoenleber et al., 2011). Results indicate that this relationship is also present when using an operationalisation of aggression where aggression is both non-physical (i.e. a noise blast does not cause direct physical pain in the same way an electric shock does) and delayed (i.e. the noise blast is delivered later, contingent on winning), a distinction which is meaningful to an attempt to understand behaviour that is dependent both on context

and the methods used to study it. The relationship between aggression and grandiose narcissism therefore does not seem to be exclusively contingent on a heightened sensitivity to threat. This expands the more established theoretical notion of narcissistic aggression as a response to negative feelings caused by a fragile sense of self-esteem being challenged or undermined. Rather, aggression in grandiose narcissistic individuals may involve more proactive behaviors, potentially due to a drive to pre-emptively assert control over a social group, signal qualities associated with aggression that are perceived as positive, and/or influence their social context to normalise aggressive behaviours and therefore facilitate future use of narcissistic aggression.

No effect of vulnerable narcissism on aggressive behaviour was observed. Current research suggests that vulnerable narcissism is strongly associated with lifestyle psychopathic traits, such as impulsivity and reactive aggression (Schoenleber et al., 2011), and conforms well to the assumption that narcissistic aggression is defensive in nature (Hart et al., 2017; Krizan & Johar, 2015). However, as discussed above, the validity of the studies supporting this association is compromised by their use of the HSNS, which by its nature renders any association between ratings on the HSNS and impulsive or reactive aggression rather unsurprising, and unlikely to be wholly representative of vulnerable narcissism as a construct. As a result, it is difficult to interpret the findings of the present study, as whilst the lack of any observable effect of narcissism seems unusual, it is nevertheless uncertain whether the lack of observable effect should be considered unusual given the somewhat problematic nature of past findings derived from the HSNS. These non-significant findings may be due to a true lack of association between VN and aggression following the social insult involved in the experiment, or due to a methodological limitations such as a lack of statistical power or issues of validity in the operationalisation of either narcissism or aggression.

Nevertheless, even if it is assumed that higher vulnerable narcissism is associated with aggression in response to a social insult, it is difficult to disentangle this from a

more general disposition towards aggressive behaviour in the present study, as almost all individuals who chose to deliver a noise blast in the first instance, also chose to deliver a noise blast in the second instance, and vice versa. As a result, it was not possible to examine any group of uniquely defensive aggressors, with the sample in the present study consisting either of unprovoked aggressors or non-aggressors.

Critique of the FFNI-SF

While the FFNI-SF is superior to the NPI and HSNS as a multi-facetted measure of both grandiose and vulnerable narcissism, the present study highlights that the FFNI-SF is not without flaws. The strong correlation observed between grandiose and vulnerable narcissism, and the shared correlation between both grandiose and vulnerable narcissism and sensitivity to provocations, suggests the FFNI-SF may also be unable to adequately distinguish grandiose and vulnerable narcissism in practice, and more specifically in the context of narcissistic aggression. This shared association with sensitivity to provocations may seem somewhat inconsistent with the notion that grandiose narcissism was not associated with a greater tendency towards reactive aggressive behaviour in the present study. However, given that narcissistic aggression is assumed to be highly context-specific, this association may reflect that grandiose narcissism is associated with reactive aggression, but that only certain provocations are sufficient to elicit retaliatory behaviour distinct from a more general tendency for assertive aggressive behaviour. This is supported by previous research; when asked to imagine social situations in which they were belittled by comments from another individual, grandiose narcissists more readily accept criticism of their personal qualities as being true compared to vulnerable narcissists (however, VN here was defined by the HSNS), but are less concerned or upset by this criticism (Hart et al., 2017). This could be due to grandiose narcissists having more of a tendency to have positive perceptions of traits that are typically considered negative, such as traits associated with their tendency for assertive (hyperactivating) aggressive strategies (e.g. accepting they are arrogant or belligerent, but viewing that as a positive quality).

If this is true, then it could be the case that whilst grandiose narcissism is related to sensitivity to provocations on a trait level, the particular design of the behavioural task and the social insult event involved were not sufficient to elicit retaliatory behaviour. Instead, in the present study grandiose narcissists behaved in a manner that was generally aggressive, and therefore eclipsed any retaliatory behaviour unique to vulnerable narcissism that may have been present. This may have been exacerbated by the deindividuation that was part of the experimental design, as all participants were kept hidden from other participants and anonymized in the game.

Critique of the Cyberball paradigm

The Cyberball paradigm has obvious utility as a simple method for eliciting aggression via an unambiguous social insult, that should be particularly effective for eliciting narcissistic aggression given the assumed sensitivity of narcissistic individuals to social insults. In the present study, it provided an easily modifiable paradigm that allowed a complete operationalisation of aggression that allowed the exploration of variables relevant to narcissistic aggression. However, a surprisingly infrequently mentioned limitation of Cyberball is its reliance on an ostentatious form of deception, as the participant is expected to believe they are playing a cooperative game remotely with two other participants, and that they are clearly excluded by the other two participants with no apparent justification.

There are a number of reasons to be skeptical about whether this form of deception is successful. The most significant is that it requires a certain degree of ‘smoke and mirrors’ from the experimenter, as they have to build in realistic features to the behavioural task such as latency whilst waiting for other participants to respond or connect, arbitrary loading screens informing the participants that under-the-hood processes of multiplayer networking are taking place, the notion of the experiment featuring other participants and other experimenters in other rooms (all of which must require a non-negligible amount of coordination) and so on. This is further complicated by the necessity for some participants to complete psychometric

questionnaires before the behavioural task, which given the significant variability in time taken to complete these questionnaires makes the notion of coordinated multiplayer activity across separate rooms more implausible. Although efforts were taken in the present study to make these features as realistic as possible, this form of almost theatrical deception is necessarily more art than science, and therefore it is not certain whether the experimenter in the present study gave a satisfactory performance to convince all participants. A largely unavoidable and acknowledged limitation of a sample consisting primarily of psychology students (for whom participating in experiments in return for course credits or monetary inconvenience allowances is routine) is prior familiarity with this form of deception from similar psychological experiments (Boynton et al., 2013). Particularly astute students may be aware that the remote multiplayer connectivity that was supposedly involved in the behavioural task is difficult to achieve and likely too sophisticated for an otherwise simplistic psychological experiment. Unfortunately in the present study no measure was included to assess whether participants believed the deception or not, although whether a self-report measure such as this would be reliable is unclear. Regardless, several participants did express that they were skeptical about whether or not they were really playing with real participants during the behavioural task. It may therefore be the case that the Cyberball paradigm is limited as an operationalisation of aggression due to the tenuous success of the necessary deception involved, which may prevent some participants from perceiving the social insult as genuinely threatening.

Further examination of grandiose and vulnerable narcissistic aggression

Findings indicated that grandiose narcissism was correlated with a willingness to engage in a task where winning players could (i) deliver a noise blast, and where losing players could (ii) receive a noise blast. This is coherent with the finding that grandiose narcissism is associated with aggression independent of provocation, as their behaviour suggests that individuals with higher grandiose narcissism were content to engage in this task independent of any particular motivation.

The correlation between sensitivity to provocations and willingness to engage in a task characterised by these two features is perhaps predictable given that those who are more willing to engage in an opportunity for retaliation following a provocation event should be assumed to be more likely to be sensitive to provocations at a trait level. Similarly, the association between sensitivity to provocations and the decision to deliver a noise blast at the second instance is predictable given this decision followed a provocation event. It is interesting that sensitivity to provocations was also associated with the decision to deliver a noise blast at the first instance, prior to this provocation event. It could be that trait sensitivity to provocations is associated with a greater disposition to aggression in general, or it could be that this association is attributable to the influence of narcissistic traits, given the relationship observed between grandiose narcissism and sensitivity to provocations, and the decision to deliver a noise blast at the first instance.

Findings indicated a correlation between vulnerable narcissism and an attitude that was more content with the ability for winning players to deliver a noise blast, but not for losers to receive a noise blast. This may indicate that vulnerable narcissists are eager to retaliate against a perceived social insult, but that they retain a fear of the costs of failure (i.e. further damage to fragile self-esteem following failed retaliation) and are therefore less content about the possibility of receiving a noise blast (Dickinson & Pincus, 2003; Grapsas et al., 2020).

Whilst appropriate for an exploratory study of this type, it is unfortunate that due to practical scheduling limitations the sample size featured in the present study could not reach a sufficiently large size to attempt analyses that might provide a deeper exploration of the precise relationships between variables, such as regression or partial correlation. Given small but nevertheless notable effect sizes (i.e. >0.20) were observed throughout the non-significant between-group analyses, a larger sample might lend the power to clarify instances where non-significance may indicate a Type-II error.

Conclusion

The findings of the present study indicate a general association between grandiose narcissism and aggressive behaviour in the Cyberball paradigm. The social insult that constituted the provocation featured in the present study did not seem to elicit a clear retaliatory response from individuals high in grandiose narcissism. Whilst it is not certain whether individuals high in vulnerable narcissism acted differently in response to the same provocation, the observed trait level associations between vulnerable narcissism and sensitivity to provocations that suggest vulnerable narcissism may be associated with retaliatory aggression in line with theoretical assumptions. However, this was directly observed in the present study. It may be that the relatively unambiguous nature of the threat featured in the Cyberball task made it difficult to disentangle the specific influence of narcissistic subtypes from a more general tendency to behave aggressively in situations featuring an unambiguous threat. The following study will therefore explore a different approach, introducing an element of ambiguity to threat perception that may emphasise individual differences. Nevertheless, the study successfully demonstrated a method that uses a laboratory based operationalisation of aggression that can measure behavioral changes associated with narcissism. The present study also demonstrated the importance of using a holistic definition for grandiose and vulnerable narcissism, and choosing an appropriate measure to reflect this.

References

Bernstein, M. J., & Claypool, H. M. (2012a). Social exclusion and pain sensitivity: Why exclusion sometimes hurts and sometimes numbs. *Personality & Social Psychology Bulletin*, *38*(2), 185–196. https://doi.org/10.1177/0146167211422449

Bernstein, M. J., & Claypool, H. M. (2012b). Not all social exclusions are created equal: Emotional distress following social exclusion is moderated by exclusion

paradigm. *Social Influence*, *7*(2), 113–130. https://doi.org/10.1080/15534510.2012.664326

Boden, J. M., Fergusson, D. M., & Horwood, L. J. (2007). Self-esteem and violence: Testing links between adolescent self-esteem and later hostility and violent behavior. *Social Psychiatry and Psychiatric Epidemiology*, *42*(11), 881–891. https://doi.org/10.1007/s00127-007-0251-7

Boynton, M. H., Portnoy, D. B., & Johnson, B. T. (2013). Exploring the ethics and psychological impact of deception in psychological research. *IRB*, *35*(2), 7–13.

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology*, *75*(1), 219–229. https://doi.org/10.1037/0022-3514.75.1.219

Bushman, B. J., Baumeister, R. F., Thomaes, S., Ryu, E., Begeer, S., & West, S. G. (2009). Looking Again, and Harder, for a Link Between Low Self-Esteem and Aggression. *Journal of Personality*, *77*(2), 427–446. https://doi.org/10.1111/j.1467-6494.2008.00553.x

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333. https://doi.org/10.1037/0022-3514.67.2.319

Cascio, C. N., Konrath, S. H., & Falk, E. B. (2015). Narcissists' social pain seen only in the brain. *Social Cognitive and Affective Neuroscience*, *10*(3), 335–341. https://doi.org/10.1093/scan/nsu072

Chester, D. S., & DeWall, C. N. (2016). Sound the alarm: The effect of narcissism on retaliatory aggression is moderated by dACC reactivity to rejection. *Journal of Personality*, *84*(3), 361–368. https://doi.org/10.1111/jopy.12164

Chrétien, S., Ensink, K., Descôteaux, J., & Normandin, L. (2018). *Measuring Grandiose and Vulnerable Narcissism in Adolescents*. *6*, 1–23. https://doi.org/10.6092/2282-1619/2018.6.1848

Dickinson, K. A., & Pincus, A. L. (2003). Interpersonal analysis of grandiose and vulnerable narcissism. *Journal of Personality Disorders*, *17*(3), 188–207. https://doi.org/10.1521/pedi.17.3.188.22146

Glover, N., Miller, J. D., Lynam, D. R., Crego, C., & Widiger, T. A. (2012). The five-factor narcissism inventory: A five-factor measure of narcissistic personality traits. *Journal of Personality Assessment*, *94*(5), 500–512. https://doi.org/10.1080/00223891.2012.670680

Grapsas, S., Brummelman, E., Back, M. D., & Denissen, J. J. A. (2020). The "Why" and "How" of Narcissism: A Process Model of Narcissistic Status Pursuit. *Perspectives on Psychological Science*, *15*(1), 150–172. https://doi.org/10.1177/1745691619873350

Hart, W., Adams, J., Burton, K. A., & Tortoriello, G. K. (2017). Narcissism and self-presentation: Profiling grandiose and vulnerable Narcissists' self-presentation tactic use. *Personality and Individual Differences*, *104*, 48–57. https://doi.org/10.1016/j.paid.2016.06.062

Hartgerink, C. H. J., van Beest, I., Wicherts, J. M., & Williams, K. D. (2015). The Ordinal Effects of Ostracism: A Meta-Analysis of 120 Cyberball Studies. *PLoS ONE*, *10*(5), e0127002. https://doi.org/10.1371/journal.pone.0127002

Hendin, H. M., & Cheek, J. M. (1997). Assessing hypersensitive narcissism: A reexamination of Murray's Narcism Scale. *Journal of Research in Personality*, *31*(4), 588–599. https://doi.org/10.1006/jrpe.1997.2204

Jamieson, J. P., Harkins, S. G., & Williams, K. D. (2010). Need Threat Can Motivate Performance After Ostracism. *Personality and Social Psychology Bulletin*, *36*(5), 690–702. https://doi.org/10.1177/0146167209358882

Krizan, Z., & Johar, O. (2015). Narcissistic rage revisited. *Journal of Personality and Social Psychology*, *108*(5), 784–801. https://doi.org/10.1037/pspp0000013

Lawrence, C. (2006). Measuring individual responses to aggression-triggering events: Development of the situational triggers of aggressive responses (STAR) scale. *Aggressive Behavior*, *32*(3), 241–252. https://doi.org/10.1002/ab.20122

Matthews, G., Jones, D., & Chamberlain, A. (1990). Refining the measurement of mood: The UWIST Mood Adjective Checklist. *British Journal of Psychology*, *81*. https://doi.org/10.1111/j.2044-8295.1990.tb02343.x

McCarthy, R. J., & Elson, M. (2018). A Conceptual Review of Lab-Based Aggression Paradigms. *Collabra: Psychology*, *4*(1), 4. https://doi.org/10.1525/collabra.104

Okada, R. (2010). The relationship between vulnerable narcissism and aggression in Japanese undergraduate students. *Personality and Individual Differences*, *49*(2), 113–118. https://doi.org/10.1016/j.paid.2010.03.017

Raskin, R. N., & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports*, *45*(2), 590. https://doi.org/10.2466/pr0.1979.45.2.590

Reidy, D. E., Foster, J. D., & Zeichner, A. (2010). Narcissism and unprovoked aggression. *Aggressive Behavior*, *36*(6), 414–422. https://doi.org/10.1002/ab.20356

Sacco, D. F., Wirth, J. H., Hugenberg, K., Chen, Z., & Williams, K. D. (2011). The world in black and white: Ostracism enhances the categorical perception of social information. *Journal of Experimental Social Psychology*, *47*(4), 836–842. https://doi.org/10.1016/j.jesp.2011.03.001

Schoenleber, M., Sadeh, N., & Verona, E. (2011). Parallel syndromes: Two dimensions of narcissism and the facets of psychopathic personality in criminally involved individuals. *Personality Disorders: Theory, Research, and Treatment*, *2*(2), 113–127. https://doi.org/10.1037/a0021870

Seah, S. L., & Ang, R. P. (2008). Differential correlates of reactive and proactive aggression in Asian adolescents: Relations to narcissism, anxiety, schizotypal traits, and peer relations. *Aggressive Behavior*, *34*(5), 553–562. https://doi.org/10.1002/ab.20269

Sherman, E. D., Miller, J. D., Few, L. R., Campbell, W. K., Widiger, T. A., Crego, C., & Lynam, D. R. (2015). Development of a Short Form of the Five-Factor Narcissism Inventory: The FFNI-SF. *Psychological Assessment, 27*(3), 1110–1116. https://doi.org/10.1037/pas0000100

Williams, K. D., & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. *Behavior Research Methods*, *38*(1), 174–180. https://doi.org/10.3758/bf03192765

Zeigler-Hill, V., & Jordan, C. H. (2011). Behind the Mask. In *The Handbook of Narcissism and Narcissistic Personality Disorder* (pp. 101–115). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118093108.ch9

3. Threat detection: narcissism in ambiguously threatening situations

3.1. Introduction

It is straightforward to describe the theoretical cognitive and behavioural processes of narcissistic aggression following the detection of a threat (as described in **Chapter 1**). The threat destabilizes a fragile sense of self-esteem, and causes decoherence between an external observed self-perception and an internal narcissistic self-perception. This deviates from the primary motivational goal of narcissism (maintaining coherence of the narcissistic identity and defending a fragile sense of self-esteem) and creates aversive feelings of social pain, physical stress, and/or negative emotional arousal. This results in the engagement of hyperactivating (assertive and confrontational) or deactivating (hostile and avoidant) aggressive strategies to remove the threat.

In **Chapter 2**, the previous study examined the association between trait narcissism and behavioural choices made following the detection of a threat. The threat featured in the previous study was unambiguous, overt and well-catered to the cognition of narcissism. However, in dynamic social environments threats of such an unambiguous and overt nature may be less frequent than threats that are ambiguous, as the vigilance associated with narcissism expands the scope of potentially threatening situations (Grapsas et al., 2020; Hane & Barrios, 2011; Muris et al., 2000). The primary utility of threat detection as a cognitive process is arguably to appraise these more ambiguous situations and provide an estimate of whether or not they constitute a threat, rather than the more simple form of appraisal involved in situations that are unambiguously threatening (Anderson et al., 2019; Grupe & Nitschke, 2013). Narcissism is arguably more influential as a factor during this process of threat detection in ambiguous situations. While situations containing an unambiguous threat are likely to be legitimately interpreted as threatening by most

individuals, the effect of individual differences becomes/may become more prominent when an event could be appraised as either threatening or non-threatening. That is to say, narcissism is assumed to influence the nature of the aggressive cognitive or behavioural response selected following a legitimate threat to their self-esteem, by predisposing an individual to take a more drastic aggressive response given the exaggerated consequences of this threat. However, in ambiguous situations it is also assumed to determine whether a threat to self-esteem is perceived at all. For example, a narcissistic individual may interpret a comment on their physical appearance as a deliberate insult and therefore a legitimate threat, whereas another individual may interpret the same comment as being innocuous and therefore legitimately non-threatening.

In the previous study, the unambiguous nature of the threat may have been responsible for difficulties disentangling the effect of narcissistic subtype differences. It may therefore be more appropriate to focus on studying narcissistic aggression in the context of ambiguously threatening situations instead, and consider not only how narcissism influences aggressive responses but also how narcissism influences the process of threat detection itself, and the content within ambiguous situations that may present a threat to narcissistic individuals.

As well as allowing a better insight into how narcissism shapes aggressive behaviour, studying ambiguous situations may provide insight into the nuances of subtype differences in narcissistic aggression. Whilst the previous study did not highlight any clear differences between vulnerable and grandiose narcissistic responses to an unambiguous threat (potentially due to methodological issues), it is nevertheless feasible to suggest that these differences may exist. However, the influence of these subtype differences may vary in ambiguous situations wherever concepts of vulnerability or grandiosity are more or less relevant to threat detection. Both grandiose narcissism and vulnerable narcissism should theoretically be associated with heightened threat sensitivity, although the previous study suggested grandiose narcissistic aggression is not necessarily exclusively motivated by defense. If

experimental accounts of vulnerable narcissistic aggression in terms of increased hostility and reactive aggression (Grapsas et al., 2020; Miller et al., 2017) are correct, then it may be that vulnerable narcissism is associated with a greater tendency to perceive ambiguously threatening situations as being legitimately threatening. This would also be coherent with theoretical assumptions concerning the association between vulnerable narcissism and hostile deactivating strategies, which facilitate a vicious cycle of increasing threat sensitivity (Miller et al., 2011; Shaver et al., 2011).

Further nuance in subtype differences may involve the motivational content of the threat. Grandiose and vulnerable narcissists may attribute different levels of salience to threats based on either direct threat in the sense of classic hostility (i.e. "another individual is insulting me"), or less direct threat in the sense of humiliation (i.e. "the actions of another individual made me feel negative about myself and therefore they must be punished / I must compensate for these negative feelings"; Hart et al., 2021; Kjærvik & Bushman, 2021). Similarly, whether a threat is interpreted as directly hostile or humiliating is likely to be ambiguous and subtype differences may emphasise one interpretation or the other. It is also possible that subtype differences are less influential in the interpretation of the contents of a threat (i.e. that it is either hostile or humiliating) than trait narcissism is more generally. If theoretical assumptions regarding the specific cognitive and motivational features of narcissistic aggression are correct (i.e. defensive response triggered by negative emotions linked to sudden awareness of decoherence with an exaggerated narcissistic self-perception), then it may be expected that humiliation should feature more strongly in narcissistic threat detection than hostility.

Studying how narcissism influences the perception of ambiguously threatening situations therefore promises to provide a deeper insight into the underlying cognition of narcissistic aggression. However, in terms of practical methodological considerations, the choice between studying narcissistic aggression in the context of ambiguous or unambiguous situations is not trivial. In the previous study it was challenging to successfully create an experimental operationalisation of an

unambiguous threat, as it was uncertain both whether participants found the threat convincing and if the threat was salient enough to produce negative emotions of a sufficient magnitude to elicit meaningfully aggressive behaviour. Making the threat involved ambiguous is likely to only magnify these issues, and therefore taking a similar approach to designing an experimental operationalisation of an ambiguous threat may not be appropriate. An alternate approach that avoids some of these issues is to rely more heavily on psychometric self-report measures and use these to both elicit and measure feelings of threat. Vignettes (short descriptions of scenes that the participant is asked to imagine; Tremblay & Belchevski, 2004) take the job of simulating a threatening situation away from the experimenter and give it to the participant themselves, as they are invited to put themselves in an imaginary scene of their own construction (given some instructions about what that scene should include from the vignette itself). Vignettes have been applied successfully to the study of narcissistic aggression (Hart et al., 2017), and have several benefits. As the scene described by a vignette is imaginary, it is not restricted by the practical limitations of a laboratory context, and although it is obviously significantly limited in salience due to being imaginary, there is at least no question of the situation being unconvincing as the design of the method inherently requires participants to convince themselves. Similarly, whilst relying on the imagination of participants means that no two participants will be responding to exactly the same version of a vignette, it does allow for the influence of individual differences (i.e. narcissism in particular) to be amplified, as they will direct both the construction of the situation and how the situation is perceived. Participants can be queried regarding their perception of the situation in the vignette using a psychometric questionnaire (for an example of this method from the literature, see Hyatt et al., 2018).

The present study therefore aims to investigate how narcissism influences the perception of threats in ambiguous situations, in an attempt to supplement the results of- and overcome limitations within the previous study. Beyond exploring a different experimental operationalisation of aggression, the present study also continues the validation of the FFNI (Glover et al., 2012). The previous study highlighted concerns

in the ability for the FFNI to distinguish grandiose and vulnerable forms of narcissistic aggression due to a correlation between subscales measuring grandiose and vulnerable traits, and a correlation between both subtype traits and sensitivity to provocations. The present study investigates these concerns further by introducing the McLean Screening Instrument for Borderline Personality Disorder (referred to as 'BPD' henceforth; Zanarini et al., 2003), a measure that may be useful as an alternative means of assessing threat sensitivity. Whilst the present study does not seek to examine pathological forms of narcissism that may be comparable more formally with characteristics of BPD, traits associated with BPD may less formally represent a type of threat sensitivity that is closely related to some features of vulnerable narcissistic cognition (i.e. fragility, defensiveness, neuroticism) not otherwise explored by more general measures of threat sensitivity. Comparing BPD and FFNI scores may help to elucidate the differences between subtypes in narcissistic aggression as measured by the FFNI. Finally, in the present study a single-item measure of self-esteem (Robins et al., 2001) was also included, both to verify theoretical assumptions regarding the relationship between narcissism and self-esteem that were not verified in the previous study, and to validate the usefulness of this extremely lightweight psychometric tool for future studies in this series.

In summary, the present study explores how narcissism influences threat detection when a threat is ambiguous, and to what extent subtype differences in narcissism are relevant. Threat is operationalised using vignettes describing unambiguously threatening, unambiguously non-threatening, and ambiguously threatening situations (see below for specific examples). Threat sensitivity more specifically is operationalised as the extent to which the vignette is perceived as directly hostile or humiliating is also explored, allowing the exploration of the influence of subtype differences in perceiving hostility or humiliation. Unambiguously threatening vignettes should overall be perceived as more hostile than both ambiguous and unambiguously non-threatening vignettes, in accordance with a basic assumption of construct validity. High grandiose narcissism is anticipated to be associated with a non-specific increase in perceived hostility across all vignette types (Reidy et al.,

2010), whereas high vulnerable narcissism is anticipated to be associated with a more specific increase in perceived hostility for unambiguously threatening and ambiguously threatening vignette types, consistent with (Kjærvik & Bushman, 2021). Furthermore, high vulnerable narcissism is anticipated to be associated with an increase in perceived humiliation across all vignette types, distinct from grandiose narcissism (Hart et al., 2021). The study also continues the investigation of the validity of the FFNI, and the relationship between trait narcissism, threat sensitivity, and self-esteem as represented by the psychometric tools used. Importantly, threat sensitivity here constitutes both a trait measure of threat sensitivity, and the vignette-derived measure of threat sensitivity described. Additionally, investigation of threat sensitivity is expanded with the inclusion of a measure investigating BPD traits, which the present study seeks to validate as a potential measure of forms of threat sensitivity that may be more relevant for narcissistic aggression, and in particular vulnerable narcissistic aggression.

3.2. Method

Participants

Participants were 198 individuals (42 male, 149 female, 2 non-binary, 5 preferred not to say) aged between 18 and 81 years (M = 32.3) recruited online through social media and psychology study recruitment emailing lists. Participants did not attend in-person but rather completed the study remotely over Qualtrics. The experiment was approved by the School of Psychology ethics committee (Ref: S1167).

Procedure

Participants read an information sheet and provided their informed consent, before responding to a series of questionnaires presented in a counterbalanced order. Participants were debriefed following completion of the survey, and could elect to be entered into a raffle for a £20 gift voucher.

Psychometric measures

Narcissism. Both grandiose and vulnerable narcissism were assessed using the 60-item Five-Factor Narcissism Inventory Short Form (Sherman et al., 2015). Participants responded to each statement on a 5 point Likert scale ranging from 1 (disagree strongly) to 5 (agree strongly). The questionnaire consists of two primary sub-scales containing items assessing grandiose (e.g. "Others say I brag too much, but everything I say is true") and vulnerable narcissism (e.g. "I often feel as if I need compliments from others to be sure of myself") respectively.

Threat sensitivity. Sensitivity to provocations were assessed using the 12-item provocations sub-scale of the 22-item Situational Triggers of Aggressive Responses (STAR) Scale (Lawrence, 2006). Participants responded to each statement (e.g. "Someone insults me") on a 5 point Likert scale ranging from 1 (very inaccurate) to 5 (very accurate).

Self-esteem. Self-esteem was assessed using the Single-Item Self-Esteem Scale. Participants responded to the single item ("I have high self-esteem") on a 5 point scale ranging from 1 (not very true of me) to 5 (very true of me).

Borderline personality disorder. Borderline personality disorder traits were assessed using the 10-item McLean Screening Instrument for Borderline Personality Disorder. Participants responded to each item with either "Yes" or "No". Each item assesses a different diagnostic trait of borderline personality disorder (e.g. "Have you been extremely moody?", "Have you chronically felt empty?").

Provocation vignettes

Participants viewed 12 vignettes (adapted from Tremblay & Belchevski, 2004), in which they were asked to imagine themselves in situations which featured a

provocation that was either unambiguously intentional (e.g. “As you are walking down the sidewalk you meet a group of three young teenage males. One walks into you and pushes you forcefully out of the way.”), unambiguously unintentional (e.g. “You notice some kids throwing a baseball in front of your
house. Their baseball flies through your front window breaking all the glass. The kids come
to tell you that they are sorry.”) or ambiguous (e.g. “Your roommate is using the phone. You tell him you need to make a very urgent call. He tells you that he will be done in a few minutes but keeps talking for another 30 minutes.”). The vignettes were divided evenly by these three types. Participants responded to each vignette by completing two 5 point items assessing perceived hostility (e.g. “to what extent is [the provoking agent] being hostile towards you?”) and estimated experienced humiliation (e.g. “to what extent would you feel humiliated?”). Each vignette type showed good internal consistency (see **Table 1**).

Data analysis

In order to contrast the effects of narcissistic subtype traits between subjects, participants were divided into high- and low-scoring groups for grandiose and vulnerable narcissism respectively, using a median split, which was deemed appropriate given the straightforward between-groups comparison required by the research question and the nature of the data (for a more detailed defense, see Iacobucci et al., 2015). An initial zero-order correlation matrix was generated to explore general relationships between variables. Two three-way 3 x 2 x 2 mixed-measures ANOVAs were then performed to analyze the effect of vignette type (intentional, unintentional, ambiguous), vulnerable narcissism (high, low), grandiose narcissism (high, low) for ratings of perceived hostility and humiliation respectively.

3.3. Results

Trait relationships. Correlations between trait variables are shown in **Table 2**. Vulnerable narcissism was positively correlated with STAR and BPD scores, as well as with perceptions of provocation in all vignette types and perceptions of humiliation in all vignette types, and negatively correlated with SISE scores, whereas grandiose narcissism was positively correlated with STAR and SISE scores, as well as perceptions of provocation in unintentional and ambiguous vignette types.

Scores on the STAR scale were positively correlated with BPD scores, as well as perceptions of provocation in intentional and ambiguous vignette types, and perceptions of humiliation in all vignette types. Scores on the BPD scale were negatively correlated with SISE scores, and positively correlated with perceptions of provocation in all vignette types as well as perceptions of humiliation in all vignette types.

Perceptions of provocation in ambiguous vignette types were positively correlated with perceptions of provocation in intentional and unintentional vignette types, whereas perceptions of humiliation in any vignette type positively correlated with perceptions of humiliation in other vignette types. Finally, perceptions of provocation in all vignette types were positively correlated with perceptions of humiliation in all vignette types.

Ratings of hostility.

A significant main effect of vignette type ($F(2, 187) = 953.27$, $p<.001$) was observed for ratings of perceived hostility. Pairwise comparisons indicated that ratings of perceived hostility in response to intentional provocation vignettes were significantly higher than ratings in response to ambiguous ($p<.001$) and unintentional ($p<.001$) provocation vignettes. The same effect was observed for ambiguous compared to unintentional provocation vignettes. A significant main effect of vulnerable

narcissism (F(1, 187) = 20.31, p<.001) and grandiose narcissism (F(1, 187) = 7.66, p<.01) was observed for ratings of perceived hostility. No interaction effects were observed for vignette type or narcissistic subtype.

Table 1. Descriptive statistics for psychometric measures.

		Mean	**SD**	**Minimum**	**Maximum**	**α**
FFNI Grandiose narcissism	High	110.26	14.49	92	157	.900
	Low	79.29	12.67	6	90	
FFNI Vulnerable narcissism	High	59.18	6.90	48	74	.859
	Low	40.08	6.62	12	47	
STAR Provocations		36.34	10.15	12	57	.809
SISE		2.88	1.25	1	5	*
Vignette (Hostility)		29.48	5.85	16	45	.795
Vignette (Humiliation)		21.55	6.41	12	44	.882
BPD		3.68	2.88	0	10	.805

* *Scale consisted of a single item*

Ratings of humiliation.

A significant effect of vignette type (F(2, 186) = 53.69, p<.001) was observed for ratings of experienced humiliation. Pairwise comparisons indicated that ratings of

Table 2. Zero order correlations between variables.

	FFNI (VN)	FFNI (GN)	STAR	SISE	BPD	Vignettes					
						Intentional Hostile	Unintentional Hostile	Ambiguous Hostile	Intentional Humiliation	Unintentional Humiliation	Ambiguous Humiliation
FFNI (Vulnerable)	1	.092	.512**	-.511**	.564**	.332**	.235**	.329**	.440**	.432**	.456**
FFNI (Grandiose)	.092	1	.298**	.372**	.067	.062	.240**	.230**	-.047	.095	.084
STAR Provocations	.512**	.298**	1	-.090	.371**	.248**	.128	.232**	.191**	.212**	.246**
SISE	-.511**	.372**	-.090	1	-.460**	-.134	-.057	-.050	-.243**	-.242**	-.195**
BPD	.564**	.067	.371**	-.460**	1	.146*	.223**	.199**	.241**	.340**	.297**
Vignette (Intentional - Hostile)	.332**	.062	.248**	-.134	.146*	1	.134	.569**	.259**	.288**	.331**
Vignette (Unintentional - Hostile)	.235**	.240**	.128	-.057	.223**	.134	1	.476**	.246**	.413**	.394**
Vignette (Ambiguous - Hostile)	.329**	.230**	.232**	-.050	.199**	.569**	.476**	1	.306**	.401**	.538**
Vignette (Intentional - Humiliation)	.440**	-.047	.191**	-.243**	.241**	.259**	.246**	.306**	1	.534**	.715**
Vignette (Unintentional - Humiliation)	.432**	.095	.212**	-.242**	.340**	.288**	.413**	.401**	.534**	1	.701**
Vignette (Ambiguous - Humiliation)	.456**	.084	.246**	-.195**	.297**	.331**	.394**	.538**	.715**	.701**	1

*** p<0.05, ** p<0.01**

experienced humiliation in response to ambiguous provocation vignettes were significantly higher than ratings in response to intentional (p<.01) and unintentional (p<.001) provocation vignettes. The same effect was observed for intentional compared to unintentional provocation vignettes (p<.001). A significant main effect of vulnerable narcissism (F(1, 187) = 31.63, p<.001) was observed for ratings of experienced humiliation. No main effect was observed for grandiose narcissism (F(1, 187) = .906, p = .342). No two-way interaction effects were observed for vignette type or narcissistic subtype.

A statistically significant three-way interaction was observed between vignette type, vulnerable narcissism and grandiose narcissism (F(2, 187) = 4.16, p = .016), indicating a two-way interaction that varied between levels of a third factor. However, post-hoc comparisons investigating this effect (performed using three two-way between-groups 2 x 2 ANOVA analyses investigating the effect of grandiose and vulnerable subtype groups for each vignette type) indicated that this interaction was attributable solely to the main effect of vulnerable narcissism reported above, with ratings of perceived humiliation for individuals across vignette types and in both the high and low grandiose narcissism groups being strongly influenced by their placement in the high or low vulnerable narcissism groups.

3.4. Discussion

Between-groups effects of vignette type and narcissism

The present study aimed to investigate how vulnerable and grandiose subtypes of narcissism influence threat detection, and to what extent perceptions of hostility and humiliation were more or less relevant to threat detection in either subtype. Perception of hostility was significantly higher across all vignette types for participants in the high grandiose narcissism group compared to participants in the low grandiose narcissism group. The same effect was observed between participants in the high compared to low vulnerable narcissism groups. This reveals a general

association between trait narcissism and threat sensitivity, that is consistent with the findings of the previous study, and affirms the hypothesised assumption that both grandiose and vulnerable narcissistic subtypes are grounded in shared cognitive features. That this effect is observed even for vignettes that were unambiguously non-threatening (unintentional) is consistent with recent findings associating narcissism and aggression even in contexts featuring no provocation (Kjærvik & Bushman, 2021). It is also possible that the association between high levels of both narcissistic subtypes it represents to an extent the partially confounding influence of more pathological forms of narcissism, that are characterized by more pathological and indiscriminate forms of aggression (Houlcroft et al., 2012). Pathological narcissism features both overt grandiose and vulnerable narcissistic traits, and so it is plausible that a subset of participants high in grandiose and/or vulnerable narcissism could represent a population with traits closer to pathological narcissism. However, were this the case, it might be anticipated that a two-way interaction effect should have been observed between grandiose and vulnerable narcissism group type (this is discussed further below).

Perception of humiliation was significantly higher across all vignette types between participants in the high vulnerable narcissism group compared to participants in the low vulnerable narcissism group, but unlike perception of hostility this effect was not observed between high compared to low grandiose narcissism groups. This reveals a distinct subtype effect of vulnerable narcissism, that may have not been observable in the previous study due to the focus on behavioural outcomes rather than underlying threat cognition. An association with humiliation is coherent with theoretical assumptions concerning the increased neuroticism and anxiety associated with vulnerable narcissism, as it would be expected this association would result in negative affect and feelings of shame featuring more strongly in the narrative of vulnerable narcissistic threat detection (Czarna et al., 2021; Miller et al., 2018). It is interesting to consider that a disposition towards perceiving potentially threatening situations as humiliating may support the theoretical association between vulnerable narcissistic aggression and deactivating strategies (Dickinson & Pincus, 2003;

Zajenkowski et al., 2018). Threats without directly hostile content may be more effectively addressed through strategies featuring rumination, retreat, and the increase in wariness and sensitivity towards future situations of this type, which also necessarily (implicitly) reaffirms that these situations are indeed threatening and justify an aggressive response. Alternatively, elevated feelings of negative affect and shame in threatening situations may more straightforwardly generate cognitive states that facilitate the relationship between perceived threats and directly aggressive behavioural responses (Pincus & Lukowitsky, 2010).

Interestingly, no two-way interaction effects were observed. That is to say, perception of humiliation and hostility did not vary between vignette types depending on whether participants belonged to either high or low grandiose narcissism or vulnerable narcissism groups. A three-way interaction was observed between vignette type, grandiose narcissism and vulnerable narcissism. However, post-hoc tests indicated this was likely attributable to the strong between-groups effect of vulnerable narcissism on ratings of humiliation causing a non-specific increase of the within-groups effect of vignette type, and indirectly influencing the between groups effect of grandiose narcissism. That it so say, the humiliation perceived by an individual in the high or low grandiose narcissism group was influenced very little by the direct between groups difference in grandiose narcissism, and instead strongly influenced by whether that individual was in either the high or low vulnerable narcissism group, creating a false interaction effect.

Independent of differences in narcissistic subtype, an aim of the present study was to validate the use of vignettes as a method, and explore the relationship between threat perception, hostility and humiliation (as part of understanding how narcissistic subtypes influence this relationship). Intentional vignettes were perceived as being significantly more hostile than ambiguous and unintentional vignettes, and ambiguous vignettes were perceived as being significantly more hostile than unintentional vignettes. This result is coherent with the intuitive assumption that the more a vignette is unambiguously threatening in content, the more it should be

perceived as being threatening, as well as research investigating aggression using these vignettes (Tremblay & Belchevski, 2004). This reaffirms the validity of the vignettes as an appropriate experimental operationalisation of threat, and suggests that the representation of the vignettes imagined by participants was consistent with the intended representation (i.e. in the vignette description). Ambiguous vignettes were perceived as being significantly more humiliating than intentional and unintentional vignettes, and intentional vignettes were perceived as being significantly more humiliating than unintentional vignettes. The latter result is again predictable given intuitive assumptions about threat perception and similarly reaffirms the validity of this measure. The former result may illustrate the importance of humiliation in the cognitive narrative of threat detection outside of unambiguous situations. Whilst in unambiguously threatening situations it is straightforward to rationalize a threat on the basis of perceived hostility (i.e. because they feature an unambiguously hostile act), ambiguously threatening situations do not allow this, and so the threat must be rationalized less directly based on how it made the individual feel. That is to say, the individual rationalizes that regardless of whether a directly hostile act was present (or even intended), the feelings of humiliation produced by the situation are sufficient in themselves to constitute a threat, and therefore the source of that humiliation should be addressed similarly to a more straightforward threat (i.e. through behavioural and/or cognitive strategies).

Within-groups effects of psychometric and experimental variables

The observed relationships between psychometric traits and experimental variables reinforce the between-groups effects described above. Vulnerable narcissism and grandiose narcissism were both positively correlated with sensitivity to provocations, consistent with the main effect of narcissistic subtype on perceived hostility, and supporting prior research linking narcissism with threat sensitivity (Kjærvik & Bushman, 2021) including the previous study. Vulnerable narcissism was negatively correlated with self-esteem, whereas grandiose narcissism was positively correlated with self-esteem. Narcissism is characterized more precisely by self-esteem fragility,

and contemporary theoretical approaches to narcissism typically avoid definitions of narcissism in terms of solely high or low self-esteem. However, self-esteem may be a more useful description at a subtype level, as fragility does not preclude the possibility that narcissistic vulnerability or grandiosity is associated with a certain average level of trait self-esteem, that may be expected to be lower or higher given the specific features of either subtype. Vulnerable narcissism uniquely was positively correlated with BPD traits, affirming the anticipated association with this more nuanced form of sensitivity, and supporting prior research linking vulnerable narcissism to anxiety, instability and reactive aggression (Du et al., 2021; Lobbestael et al., 2014; Miller et al., 2017). Relationships between vulnerable narcissism, grandiose narcissism and experimental variables are explored in the context of between-subjects analyses below.

Sensitivity to provocations was positively correlated with BPD traits, consistent with the assumption that threat sensitivity is a fundamental component of this conception of BPD. Sensitivity to provocations was positively correlated with perceived hostility in intentional and ambiguous vignette types. This is consistent with the intended perception of intentional and ambiguous vignettes as being potentially threatening (distinct from unintentional vignettes which should be perceived as non-threatening at baseline), and so individuals who are more sensitive to threats should naturally perceive these vignette types as more directly hostile. Sensitivity to provocations was also positively correlated with feelings of humiliation in all vignette types, which may suggest that individuals who are more sensitive to threats are also more likely to attribute negative personal significance to threats and associate them with the experience of negative affect (Corr, 2004).

BPD traits were negatively correlated with self-esteem, which is coherent both with the anxious and depressive component of BPD (Jayaro et al., 2011), and with the association between BPD traits and vulnerable narcissism. BPD traits were also positively correlated with perceived hostility in all vignette types, and feelings of humiliation in all vignette types. As described above in the context of sensitivity to

provocations, this relationship is consistent with the nature of these vignette types, but the extension of attributing hostile intent to unintentional vignette types may reflect the more pathological nature of threat sensitivity in the context of BPD (Martino et al., 2015).

Critique of experimental design

In addition to investigating subtype differences in narcissistic threat detection, the present study also aimed to continue the exploration of successful methods for studying narcissism and aggression. To continue the validation of the FFNI, a concern regarding the FFNI highlighted in the previous study was the high correlation observed between grandiose and vulnerable narcissism, and given their shared correlation with sensitivity to provocations, there was no way to discern them from one another either directly or indirectly at a trait level. In the present study, however, little to no correlation was observed between grandiose narcissism and vulnerable narcissism as measured by the FFNI, suggesting the FFNI may be an effective measure of subtype traits. Given the significantly larger sample size featured in the present study, it may be more likely that the concerning result highlighted in the previous study was attributable to limited statistical power.

Whilst the shared association between grandiose and vulnerable narcissism and sensitivity to provocations remained, only vulnerable narcissism was correlated with the newly introduced BPD measure, highlighting a trait measure of sensitivity that more uniquely reflects the cognition of vulnerable narcissistic aggression (i.e. neuroticism, negative affect, etc.), distinct from the type of threat sensitivity associated with grandiose narcissistic aggression or narcissism more generally. Finally, the SISE measure introduced in the present study seemed to function effectively as a lightweight measure of self-esteem insofar as intuitive relationships were observed between SISE scores and the other psychometric measures used. This provides encouraging support of the usefulness of a BPD trait measure and SISE in the study of narcissistic aggression, and also further supports the validity of the FFNI

given its theoretically coherent relationships with these measures demonstrated in the present study.

As described above, the results of the present study suggest the vignettes performed their role successfully as an operationalisation of threat in the context of narcissistic aggression, insofar as participant perceptions of each vignette type were consistent with the intended perception, and several interesting and theoretically coherent relationships emerged between vignette perceptions and narcissistic traits. The lack of any specific limitations highlighted in the present study, however, does not militate a more fundamental limitation inherent to vignettes, specifically that the threatening situations they create are completely simulated and within the control of the participant. This is limitation is a double-edged sword: as described above, the simulated nature of vignettes removes the practical limitations of other operationalisations of threat, and amplifies individual differences. Nevertheless, arguably an important feature of threatening situations as they occur in more naturalistic contexts is that they represent challenges to the control an individual has over their environment, and therefore require (and justify) behavioral and cognitive strategies meant to remove the threat and regain control (Greenaway et al., 2015; Rapee, 1997). Without this challenge to individual control, it is uncertain to what extent vignettes are able to produce genuine simulations of threat.

Conclusions

The results of the present study indicate that threatening situations that are both unambiguous and ambiguous in their content are perceived as being more directly hostile in nature by individuals high in either grandiose narcissism or vulnerable narcissism. This helps to develop the account of narcissistic aggressive behaviour provided by the previous study, wherein only the influence of grandiose narcissism was evident in behaviour. Shifting to an account in terms of threat perception better elucidates the influence of both subtypes, highlighting that although behaviour may manifest differently, the basic features of underlying cognition may be shared. For

example, vulnerable narcissism may be characterised by a similar core threat sensitivity as grandiose narcissism, but respond to threats through less overt forms of behaviour. Importantly, as well as identifying subtype similarities, subtype differences in threat perception are also illuminated, with individuals high in vulnerable narcissism uniquely perceiving threatening situations as being more humiliating. This is coherent with the hypothesized association between vulnerable narcissism and aggressive cognition characterised more by neuroticism and negative affect, where feelings of humiliation may feature more prominently in justifying narcissistic aggression.

The results of the present study also demonstrate the importance of considering the ambiguity of a threatening situation as a factor that may influence threat perception. Whilst unambiguously intentional threats were seen as the most hostile, ambiguous threats were seen as the most humiliating. This suggests feelings of humiliation may provide a justification for feelings of threat in situations where these feelings cannot otherwise be justified by outright perceived hostility. The ambiguity of a threatening situation may therefore strongly influence the underlying cognition of threat perception. In particular, given the association between vulnerable narcissism and perceived humiliation, ambiguously threatening situations may more readily facilitate vulnerable narcissistic threat sensitivity. This in turn would suggest that vulnerable narcissistic responses to threat would be more pronounced in ambiguously threatening situations, which may explain why no clear vulnerable narcissistic behaviour was observed in response to the unambiguously threatening stimulus featured in the previous study.

Additionally, but no less significantly, several important methodological conclusions are made, regarding the validity of the FFNI as a metric of narcissism, the potential utility of BPD traits as an operationalisation of more distinctly narcissistic forms of aggressive cognition, and the success of vignettes as an operationalisation of threat. The present study therefore advances the understanding of the cognition of narcissistic aggression, through identifying both differences and similarities between

narcissistic subtypes in processes of threat detection, and facilitates future research into this topic through the validation of several useful psychometric methods.

References

Anderson, E. C., Carleton, R. N., Diefenbach, M., & Han, P. K. J. (2019). The Relationship Between Uncertainty and Affect. *Frontiers in Psychology*, *10*, 2504. https://doi.org/10.3389/fpsyg.2019.02504

Corr, P. J. (2004). Reinforcement sensitivity theory and personality. *Neuroscience & Biobehavioral Reviews*, *28*(3), 317–332. https://doi.org/10.1016/j.neubiorev.2004.01.005

Czarna, A. Z., Zajenkowski, M., Maciantowicz, O., & Szymaniak, K. (2021). The relationship of narcissism with tendency to react with anger and hostility: The roles of neuroticism and emotion regulation ability. *Current Psychology*, *40*(11), 5499–5514. https://doi.org/10.1007/s12144-019-00504-6

Dickinson, K. A., & Pincus, A. L. (2003). Interpersonal analysis of grandiose and vulnerable narcissism. *Journal of Personality Disorders*, *17*(3), 188–207. https://doi.org/10.1521/pedi.17.3.188.22146

Du, T. V., Miller, J. D., & Lynam, D. R. (2021). The relation between narcissism and aggression: A meta-analysis. *Journal of Personality*, *n/a*(n/a). https://doi.org/10.1111/jopy.12684

Glover, N., Miller, J. D., Lynam, D. R., Crego, C., & Widiger, T. A. (2012). The five-factor narcissism inventory: A five-factor measure of narcissistic personality traits. *Journal of Personality Assessment*, *94*(5), 500–512. https://doi.org/10.1080/00223891.2012.670680

Grapsas, S., Brummelman, E., Back, M. D., & Denissen, J. J. A. (2020). The "Why" and "How" of Narcissism: A Process Model of Narcissistic Status Pursuit. *Perspectives on Psychological Science*, *15*(1), 150–172. https://doi.org/10.1177/1745691619873350

Greenaway, K. H., Storrs, K. R., Philipp, M. C., Louis, W. R., Hornsey, M. J., & Vohs, K. D. (2015). Loss of control stimulates approach motivation. *Journal of*

Experimental Social Psychology, *56*, 235–241. https://doi.org/10.1016/j.jesp.2014.10.009

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nature Reviews. Neuroscience*, *14*(7), 488–501. https://doi.org/10.1038/nrn3524

Hane, A. A., & Barrios, E. S. (2011). Mother and child interpretations of threat in ambiguous situations: Relations with child anxiety and autonomic responding. *Journal of Family Psychology*, *25*(5), 644–652. https://doi.org/10.1037/a0024149

Hart, W., Adams, J. M., & Tortoriello, G. (2017). Narcissistic responses to provocation: An examination of the rage and threatened-egotism accounts. *Personality and Individual Differences*, *106*, 152–156. https://doi.org/10.1016/j.paid.2016.10.049

Hart, W., Tortoriello, G. K., & Richardson, K. (2021). Provoked Narcissistic Aggression: Examining the Role of De-Escalated and Escalated Provocations. *Journal of Interpersonal Violence*, *36*(9–10), 4832–4853. https://doi.org/10.1177/0886260518789901

Houlcroft, L., Bore, M., & Munro, D. (2012). Three faces of Narcissism. *Personality and Individual Differences*, *53*(3), 274–278. https://doi.org/10.1016/j.paid.2012.03.036

Hyatt, C. S., Sleep, C. E., Lamkin, J., Maples-Keller, J. L., Sedikides, C., Campbell, W. K., & Miller, J. D. (2018). Narcissism and self-esteem: A nomological network analysis. *PLoS ONE*, *13*(8), e0201088. https://doi.org/10.1371/journal.pone.0201088

Iacobucci, D., Posavac, S., Kardes, F., Schneider, M., & Popovich, D. (2015). The median split: Robust, refined, and revived ☆. *Journal of Consumer Psychology*, *25*, 690–704.

Jayaro, C., De La Vega, I., Bayon-Palomino, C., Díaz-Marsá, M., Montes, A., Tajima, K., López-Ibor, J. J., & Carrasco, J. L. (2011). Depressive-type emotional response pattern in impulsive-aggressive patients with borderline personality disorder. *Journal of Affective Disorders*, *135*(1–3), 37–42. https://doi.org/10.1016/j.jad.2011.06.040

Kjærvik, S. L., & Bushman, B. J. (2021). The link between narcissism and aggression: A meta-analytic review. *Psychological Bulletin, 147*(5), 477–503. https://doi.org/10.1037/bul0000323

Lawrence, C. (2006). Measuring individual responses to aggression-triggering events: Development of the situational triggers of aggressive responses (STAR) scale. *Aggressive Behavior, 32*(3), 241–252. https://doi.org/10.1002/ab.20122

Lobbestael, J., Baumeister, R. F., Fiebig, T., & Eckel, L. A. (2014). The role of grandiose and vulnerable narcissism in self-reported and laboratory aggression and testosterone reactivity. *Personality and Individual Differences, 69*, 22–27. https://doi.org/10.1016/j.paid.2014.05.007

Martino, F., Caselli, G., Berardi, D., Fiore, F., Marino, E., Menchetti, M., Prunetti, E., Ruggiero, G. M., Sasdelli, A., Selby, E., & Sassaroli, S. (2015). Anger rumination and aggressive behaviour in borderline personality disorder. *Personality and Mental Health, 9*(4), 277–287. https://doi.org/10.1002/pmh.1310

Miller, J. D., Hoffman, B. J., Gaughan, E. T., Gentile, B., Maples, J., & Keith Campbell, W. (2011). Grandiose and vulnerable narcissism: A nomological network analysis. *Journal of Personality, 79*(5), 1013–1042. https://doi.org/10.1111/j.1467-6494.2010.00711.x

Miller, J. D., Lynam, D. R., Hyatt, C. S., & Campbell, W. K. (2017). Controversies in Narcissism. *Annual Review of Clinical Psychology, 13*, 291–315. https://doi.org/10.1146/annurev-clinpsy-032816-045244

Miller, J. D., Lynam, D. R., Vize, C., Crowe, M., Sleep, C., Maples-Keller, J. L., Few, L. R., & Campbell, W. K. (2018). Vulnerable Narcissism Is (Mostly) a Disorder of Neuroticism. *Journal of Personality, 86*(2), 186–199. https://doi.org/10.1111/jopy.12303

Muris, P., Luermans, J., Merckelbach, H., & Mayer, B. (2000). "Danger is lurking everywhere". The relation between anxiety and threat perception abnormalities in normal children. *Journal of Behavior Therapy and Experimental Psychiatry, 31*(2), 123–136. https://doi.org/10.1016/S0005-7916(00)00016-1

Pincus, A. L., & Lukowitsky, M. R. (2010). Pathological narcissism and narcissistic personality disorder. *Annual Review of Clinical Psychology, 6*, 421–446. https://doi.org/10.1146/annurev.clinpsy.121208.131215

Rapee, R. M. (1997). Perceived threat and perceived control as predictors of the degree of fear in physical and social situations. *Journal of Anxiety Disorders*, *11*(5), 455–461. https://doi.org/10.1016/s0887-6185(97)00022-4

Reidy, D. E., Foster, J. D., & Zeichner, A. (2010). Narcissism and unprovoked aggression. *Aggressive Behavior*, *36*(6), 414–422. https://doi.org/10.1002/ab.20356

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *27*(2), 151–161. https://doi.org/10.1177/0146167201272002

Sherman, E. D., Miller, J. D., Few, L. R., Campbell, W. K., Widiger, T. A., Crego, C., & Lynam, D. R. (2015). Development of a Short Form of the Five-Factor Narcissism Inventory: The FFNI-SF. *Psychological Assessment*, *27*(3), 1110–1116. https://doi.org/10.1037/pas0000100

Tremblay, P. F., & Belchevski, M. (2004). Did the Instigator Intend to Provoke? A Key Moderator in the Relation Between Trait Aggression and Aggressive Behavior. *Aggressive Behavior*, *30*(5), 409–424. https://doi.org/10.1002/ab.20027

Zajenkowski, M., Maciantowicz, O., Szymaniak, K., & Urban, P. (2018). Vulnerable and Grandiose Narcissism Are Differentially Associated With Ability and Trait Emotional Intelligence. *Frontiers in Psychology*, *9*, 1606. https://doi.org/10.3389/fpsyg.2018.01606

Zanarini, M. C., Vujanovic, A. A., Parachini, E. A., Boulanger, J. L., Frankenburg, F. R., & Hennen, J. (2003). A screening measure for BPD: The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD). *Journal of Personality Disorders*, *17*(6), 568–573. https://doi.org/10.1521/pedi.17.6.568.25355

4. Conflict monitoring as narcissistic threat sensitivity

4.1. Introduction

The studies presented in previous chapters so far demonstrate that narcissistic aggression can to some extent be successfully studied using a methodological approach relying on conventional behavioural and individual differences measures. However, the limitation of operationalisations of aggression that rely fundamentally on self-report assessment is that they will always reflect after-the-fact perceptions of aggression that are heavily biased by individual beliefs. As discussed in the context of the previous study, the effects of these individual biases are not necessarily confounding, as these biases ultimately constitute the forces that drive narcissistic aggression through influencing the perception of threatening situations. Nevertheless, whilst the influence of these biases upon aggressive behaviour is desirable, it is not desirable if these biases also influence the measurement of this behaviour itself. Otherwise, it is unclear to what extent one is measuring narcissistic aggression versus narcissistic self-perception – in other words, the difference measuring actual behaviour versus measuring perception of behaviour.

A neuroscience-informed approach to the development of narcissistic theory

With conventional methodological approaches being limited in this way, it may be useful to take an alternative approach to understanding the relationship between narcissism and aggression. The primary utility of an alternative approach allowing a more objective assessment of narcissistic aggression is expanding theoretical descriptions of narcissism to incorporate ideas that are not derived from how narcissists see themselves or how narcissists are seen by others. Furthermore, whilst the arbitrary reduction of narcissism into more concrete cognitive or even physiological terms is not useful, careful and deliberate reduction may help identify

and trim unnecessary complexity when approaching a phenomenon as apparently nuanced and multifaceted as narcissistic personality. The significant increase in the availability of neuroimaging literature in the last decade makes insights from neuroscience a useful resource for considering what this alternative approach might involve. To date, neuroscientific data has had little influence in shaping the conception of narcissism and narcissistic aggression (Zarnowski et al., 2021), and neuroimaging studies investigating narcissism in non-clinical populations are scarce. The few available studies highlight that narcissistic cognition is focused almost entirely in neocortical areas associated with diverse associative and executive functions (Cascio et al. 2015; Chester & DeWall 2016; Fan et al. 2011; Jauk et al. 2017). Research suggests network activity in these areas with activity in these areas is associated with the experience of social pain and rejection in narcissistic individuals, with social pain being the acute stressor initiating negative affect and aggression in the absence of more 'concrete' stressors (Burklund et al. 2007; Eisenberger et al. 2003; Kawamoto et al. 2012). Reactivity in the dorsal anterior cingulate cortex (dACC) specifically reportedly directly moderates the relationship between narcissism and aggression in a threatening situation (Chester & DeWall 2016). The presence of the dACC as a common feature across neuroimaging studies investigating narcissism and its involvement as an apparently important mediator in narcissistic aggression is coherent with the strong involvement of this region in social pain more generally (Eisenberger, 2015; Rotge et al., 2015).

Conflict detection as a fundamental description of narcissistic aggression

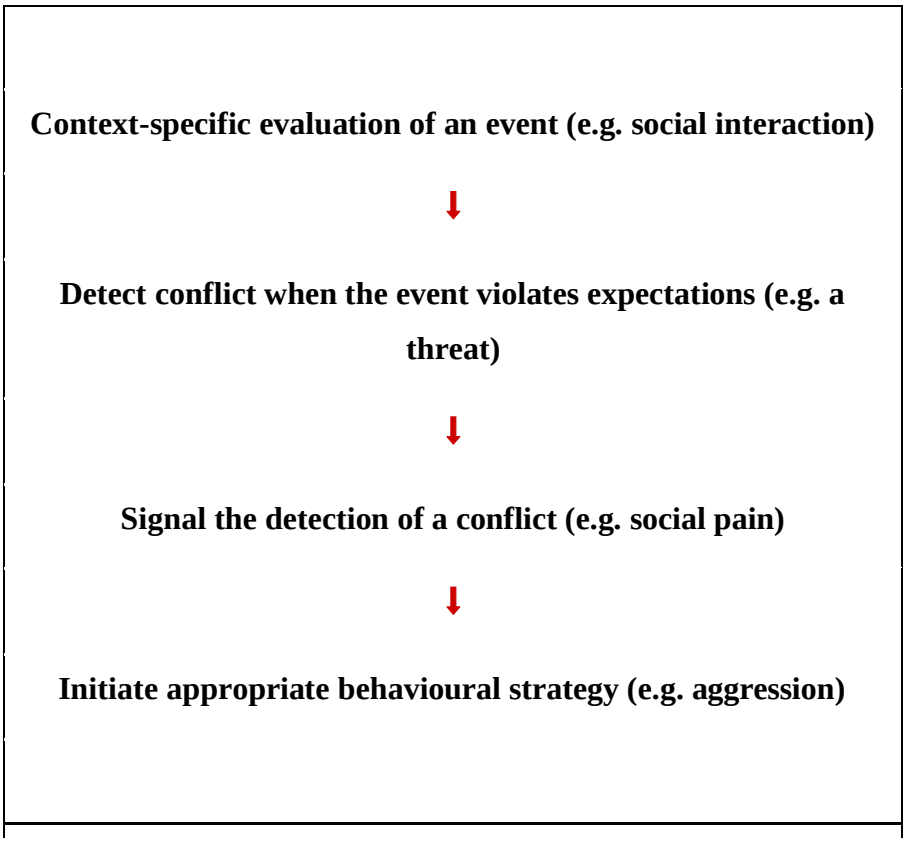
The fundamental theoretical suggestion defining the investigations described in the present study is that understanding how the deeper function of the dACC relates to its involvement in social pain may provide a theory that describes social pain, rejection sensitivity, and narcissistic aggression in terms of the same fundamental cognitive process. The answer may lie in the other, more well-established function attributed to the dACC; conflict monitoring and error detection (Botvinick et al. 2004). Conflict monitoring is a cognitive process whereby inconsistencies are spotted in received

information. These inconsistencies are generally defined by instances where an outcome observed in received information violates the outcome that was expected based on prediction. Error detection is a similar process, involved in spotting when an expected correct outcome is violated by an observed incorrect outcome. Whilst error detection is not by necessity associated with the same cognitive process as conflict monitoring, there is evidence to suggest that the dACC is the source of the error-related negativity (an event-related potential observed following an individual making an error), which suggests they may be closely related processes within the dACC (Orr & Hester 2012). Other neuroimaging studies investigating the function of the dACC support this description, with numerous studies finding that the dACC is strongly associated with context-specific information processing, connecting numerous input pathways with action-related outputs (Heilbronner & Hayden 2016). This context-specific evaluation would be central to conflict monitoring, as whether or not an event constitutes a conflict is necessarily dependent on the situation in which it occurs.

Whilst this shared association with both narcissistic aggression and conflict monitoring within the dACC may intuitively seem unrelated, there are interesting functional similarities that emerge upon closer inspection. If narcissistic aggression can be fundamentally understood as an exaggerated response to a threat in the form of social pain inflicted upon them, then it should be associated with a system wherein incoming information is monitored for a threat, a social pain signal is generated to indicate the detection (and possibly salience) of this threat, and an appropriate behavioural response is initiated. This closely aligns with the system described above, wherein the dACC performs context-specific evaluation of incoming information, detection of conflicts, and feeds appropriate information forward into action outputs. There is no necessary distinction between a conflict (the latter) and a threat (the former), as both consist of the same process wherein an expected outcome (e.g. no threat event present) is violated by an observed outcome (e.g. threat event present). Similarly, there is no necessary distinction between detecting a conflict and initiating

an action, and generating a social pain signal and initiating a behavioural response; this is simply describing the same process at two different levels of abstraction.

Figure 1. A hypothetical mechanism for narcissistic aggression in terms of dACC processing.



Altered functioning in the dACC may therefore be associated with narcissistic aggression via this system, wherein some change in dACC activity results in either (**1)** conflicts being more readily detected, **(2)** the generated alarm signal being more salient, **(3)** the action output being more inclined towards an aggressive response, or (**4)** a combination of the above. Regardless of which assumption may be the case, this hypothetical system presents a new mechanism for narcissistic aggression in terms of conflict detection (see **Figure 1**). If this mechanism is accurate, it may provide a methodological option that substantially reduces the difficulty of measuring narcissistic aggression (and, to some extent, narcissism), which in turn may simplify

the process of understanding the phenomenon in more detail (e.g. further understanding of subtype differences and underlying cognitive mechanisms). Further predictions relating more specifically to narcissistic traits are made below in the overview of the present study.

Aims and objectives of the present study

The present study encompasses two experiments examining the first assumption described above, specifically whether narcissistic aggression can be measured as an increased sensitivity to conflict detection. To investigate this, in each experiment two distinct but complementary behavioural measures of conflict detection are used. The use of two measures instead of a single measure allows for a certain amount of redundancy, given the exploratory nature of the study and the lack of prior literature that might inform the methods used, and allows the effectiveness of either method to be better evaluated in light of the other. Furthermore, as a persistent limitation noted in the previous studies and the broader literature is the inability to confidently know how different operationalisations of the same concept relate to one another, the use of two methods allows for a comparison that may highlight important inconsistencies or similarities between methods. These may then be relevant to future research studying the relationship between aggression and conflict detection.

The first measure featured in both experiments is an adapted version of the Eriksen flanker task (Eriksen & Eriksen 1974), which in the first experiment features letter-based stimuli and in the second experiment features face-based stimuli (explained below). A flanker task requires participants to respond to a target stimulus whilst ignoring flanking stimuli that are either congruent or incongruent with the target stimuli. In order to perform the task, a participant has to actively monitor their behaviour to detect conflicts and respond accordingly. This measure therefore provides both a means of assessing conflict monitoring abilities, and putting participants in a situation in which they encounter conflicts. However, as this task assesses a very simple and unambiguous conflict, its use is primarily in investigating

the limits to which the relationship between conflict detection and narcissistic aggression can be reduced. That is to say, it provides insight into whether narcissism influences the most fundamental forms of conflict detection.

This insight is important, but highlights an equally important consideration: whilst there is a plausible link between narcissistic aggression and conflict detection, it is unclear whether this is reducible to a general association (i.e. narcissistic individuals are fundamentally more sensitive to all forms of cognitive conflict, leading to an increase in threat sensitivity that only appears specific), or whether narcissism only influences conflict detection in certain contexts (e.g. the social contexts in which narcissism primarily exerts its influence, leading to a legitimately specific increase in threat sensitivity). This consideration is supported by the previous study which demonstrated the context-specific influence of narcissism, and in particular the relevance of narcissism in ambiguously threatening situations. An adapted version of the task introducing these ambiguous and more socially relevant elements may therefore provide a more naturalistic insight into narcissistic conflict detection in the event that it cannot be reduced to basic cognition, and may more generally highlight any nuances missed by a more reductive approach to measuring narcissistic conflict detection. In order to investigate this, the Eriksen flanker task featured in the second experiment is adapted to feature a set of stimuli consisting of emotional faces. Face-based flanker tasks have seen some use in the literature (Fenske & Eastwood, 2003; Tannert & Rothermund, 2020). Interpreting facial emotions, and responding to sudden conflicting changes in facial emotions, may be suitably representative of the kind of conflict detection that is relevant in the social contexts where narcissism is most influential (De Panfilis et al., 2019).

As in the classic version of the task, flanking and target stimuli may be either congruent or incongruent, with stimuli being categorized by the type of emotion they are displaying. However, the face-based flanker task featured in the second experiment introduces an additional nuance through the use of three types of emotional face stimuli; happy, angry and neutral. Happy and angry emotion faces are

taken to represent unambiguously non-threatening and threatening stimuli respectively, allowing the influence of these factors on the association between narcissistic aggression and conflict detection to be examined. As narcissistic threat sensitivity should be most relevant in response to threatening stimuli, task performance should be enhanced when identifying angry emotion faces, and reduced when ignoring angry emotion faces as incongruent flanking stimuli. Introducing a third and less distinct emotional face displaying 'neutral' features is an exploratory decision based on the findings of the previous empirical chapter (see **Chapter 3**), which suggested ambiguity may be influential in narcissistic aggression. Specifically, narcissism may influence the interpretation of ambiguous (neutral) facial emotions to perceive them as being more hostile. The result of this would be that neutral emotion faces would appear congruent with unambiguously angry emotion faces, increasing task performance when identifying angry emotion faces by effectively increasing the number of congruent trials featuring angry emotion faces (and therefore making the task easier). Conversely, this would also result in decreased task performance when identifying neutral emotion faces, as these would be incorrectly perceived as angry emotion faces more frequently. This face-based flanker task therefore provides an assessment of the extent to which the association between narcissistic aggression and conflict detection is sensitive to the presence more explicitly threatening and 'socially-relevant' stimuli, and broader insight into the influence of narcissistic biases on conflict detection involving ambiguous stimuli.

The second measure featured in both experiments, which relies on the ability of the flanker task to present conflicting stimuli, is changes in pupil dilation amplitude related to trials on the flanker task. Pupil dilation is associated with numerous events, but is a well-established index of increases in cognitive control (van der Wel & van Steenbergen 2018), which in turn is considered a fundamental response to conflict detection (Botvinick et al. 2004). As a result, when examined immediately following trials on the flanker task, and in particular trials in which an incongruent (or additionally, in the case of the face-based flanker task, hostile) flanking stimuli is presented, pupil dilation (i.e. increase in pupil size) may act as a measure of conflict

detection, and may allow conflict monitoring to be assessed independently of actual task performance (i.e. the extent to which task-related conflicts influence judgment), if task performance does not provide a sufficiently robust measure by itself. The use of pupil dilation as a measure in personality psychology research has been limited but there are several examples of its successful use, indicating that its use as a trait-related measure in the present study is plausible (Al-Samarraie et al., 2018; Hubert Lyall & Järvikivi, 2021; Inzlicht et al., 2015; Sleegers et al., 2017; Unsworth et al., 2019). With particular reference to the face-based flanker task featured in the second experiment, pupil dilation has been shown to be associated with viewing affective images, further suggesting it should be suitable for use as a complementary measure (Snowden et al., 2016). In the context of narcissistic aggression more specifically, Sleegers et al. (2017) observed task-related changes to pupil size in response to ostracism (using a Cyberball task) as a psychophysiological response to social pain in response to an ego-threat. That is to say, social threat was associated with a pupillary response caused by the experience of social pain, which features prominently in the cognition of narcissistic aggression, and arguably represents a form of conflict detection. Notably, this study further reaffirms the choice to conduct two experiments investigating both basic conflict detection and conflict detection in a socially-relevant task, highlighting the potential importance of socially-relevant factors in pupillary responses.

The two experiments documented in the present study therefore investigate the hypothesised association between narcissistic aggression and conflict detection by examining the association between performance on a conflict detection task, task-related changes in pupil dilation associated with conflict detection, and trait measures of narcissism. Importantly, the second experiment introduces a conflict detection task containing ambiguous stimuli that might reflect narcissistic biases in conflict detection.

However, absent from this hypothesized mechanism of narcissistic aggression is the influence of subtype differences, which have been demonstrated to act as influential

factors in narcissistic aggression, both research literature (Hart et al., 2017) and in the previous studies documented in the thesis. Indeed, the theoretical framework of narcissistic aggression informing the thesis (see **Chapter 1**) and the results of previous studies provide sufficient insight to make several other predictions. Considering this link between narcissistic aggression and threat detection more specifically in the context of narcissistic traits, the distinctions between vulnerable and grandiose subtypes highlighted by the previous studies suggests subtype differences may be expected to emerge in the strategy associated with response to a conflict. Given the association between vulnerable narcissism and hypersensitivity, neuroticism, and anxiety (Houlcroft et al. 2012; Rogoza et al. 2018), and the suggested association with vulnerable narcissism and feelings of humiliation and 'deactivating' aggressive strategies discussed in the previous studies, it is possible that the vulnerable subtype of narcissistic aggression may be associated with behaviour that reflects this. Pupil dilation is associated with noradrenergic signalling, which itself is closely associated with activating 'flight or fight' pathways in situations where an alarming event requires an immediate stereotyped response (Larsen & Waters 2018). As a result, trial-related pupil dilation response may also be part of a stereotyped panic response associated with the alarm signal generated when a conflict is detected. This alarm response may be more salient for the vulnerable subtype of narcissistic aggression, potentially amplifying any trial-related changes in pupil dilation associated with purely cognitive aspects of conflict monitoring, resulting in the enhanced effect observed. This is further supported by the association between vulnerable narcissism and avoidance behaviours; vulnerable narcissists may favour a deactivating strategy to disengage from the conflict as rapidly as possible and protect a narcissistic self-image (Hart et al. 2017). It may therefore be the case that vulnerable narcissism is associated with changes in pupil dilation of a greater magnitude than grandiose narcissism.

In summary, the following two experiments that constitute the present study investigate a hypothesized link between narcissistic aggression and conflict monitoring. It is predicted that individuals high in narcissism will exhibit increased

conflict detection sensitivity. In the first experiment, this is represented as increased performance on a conventional flanker task. In the second experiment, this is similarly represented as overall enhanced performance on a face-based flanker task, but due to the introduction of affective stimuli the association between narcissistic aggression and conflict detection is predicted to be more nuanced. It is predicted that narcissistic biases influencing the perception of ambiguous stimuli should result in reduced performance when neutral emotion faces are target stimuli. It is also predicted that attentional bias towards threatening stimuli should result in enhanced performance when angry emotion faces are target stimuli, and reduced performance when angry emotion faces are flanking stimuli. In both experiments, the hypothesized association between narcissistic aggression and conflict detection is predicted to be observed as enhanced increases in pupil dilation following perceived conflicts. Furthermore, individuals high in vulnerable narcissism are predicted to exhibit the above effects to a greater magnitude relative to individuals high in grandiose narcissism. Finally, in order to verify the hypothesis that pupil dilation indicates enhanced conflict sensitivity and related increases in cognitive control, there should be a trait-independent association between trial-related changes in pupil dilation and flanker task performance in both experiments.

4.2. Experiment 1

4.2.i. Method

Participants

Participants were 95 students (20 male, 75 female) aged between 18 and 31 years (M = 21.59, SD = 3.09) attending the University of Nottingham. Participants were either Psychology students taking part in order to earn research credits, or students recruited from advertisements distributed via email. Participants were told they would be taking part in a study investigating how eye-tracking and personality traits could be

used to predict performance on cognitive tasks. The experiment was approved by the School of Psychology ethics committee (Ref: S1101).

Procedure

Participants were asked to fill out the Five Factor Narcissism Inventory (Short Form; Sherman et al. 2015), the Single-Item Self-Esteem Questionnaire (Robins et al. 2001), the Buss-Perry Aggression Questionnaire (Buss & Perry 1992) and the McLean Screening Instrument for Borderline Personality Disorder (Zanarini et al. 2003), at either the start or the end of the study counterbalanced across participants. Participants then completed the experimental task using a computer running Experiment Builder (version 2.1.140; SR Research 2017).

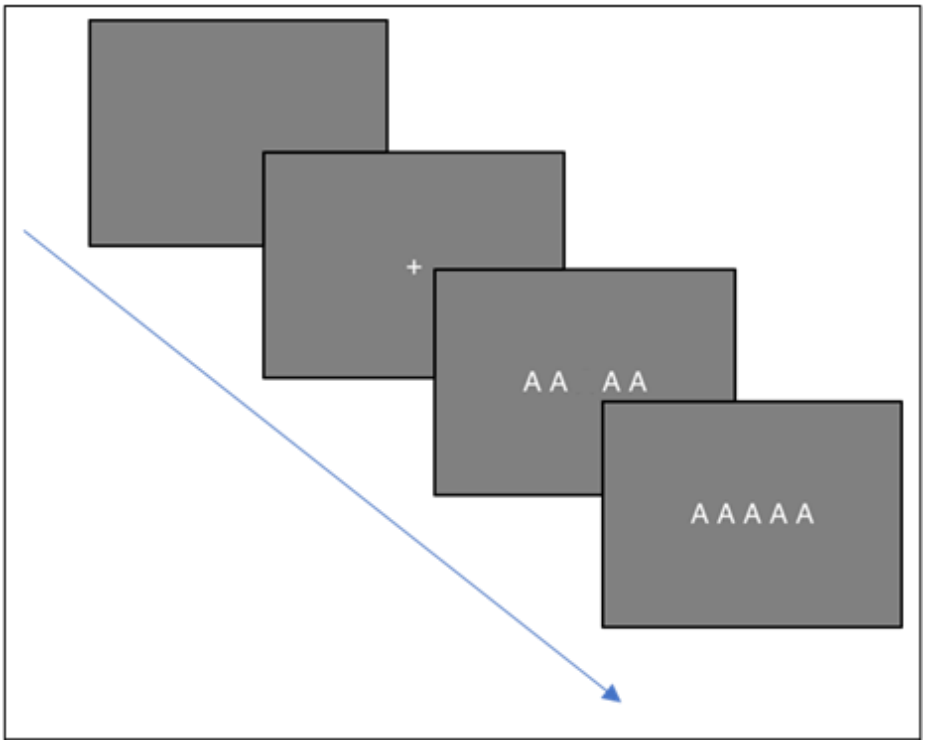
Eye-tracking specifications

Pupil diameter was measured throughout the experiment using EyeLink 1000 (SR Research 2017) eye-tracking hardware. The EyeLink 1000 is a desk-mounted device located in front of the monitor on which the participant performs the experimental task. Participants were asked to position their head comfortably on a chin rest for the duration of the experiment and to avoid moving their head during the experiment as much as possible, to ensure movement-related error was kept minimal. Pupil diameter is influenced by several factors beyond conflict-related changes in cognitive control (the dependent variable in pupillometry data), such as ambient light level, stimulus luminosity, and ambient temperature, which were kept as constant as possible throughout the experiment (Holmqvist et al. 2011). Furthermore, as described above, the latency between phasic changes in pupil dilation necessitated an inter-trial interval (i.e. blank inter-trial period and fixation period) to accommodate both a rise to peak pupil dilation and return to baseline. Other factors such as state anxiety, wakefulness and the use of substances such as caffeine are believed to influence baseline pupil diameter, which are difficult to meaningfully control for within the context of the experiment. Instead, these factors are controlled for in the present study by using

relative rather than absolute measures of pupil dilation and blink rate. Before the beginning of the task, the EyeLink 1000 was calibrated to ensure an average error of no higher than 0.5 degrees using a nine-point grid.

Behavioural task

Figure 2. Diagram showing the phases of a congruent trial on the flanker task.



Flanker task. Participants were required to respond correctly to one of two possible target letters by pressing the corresponding letter on a keyboard, whilst ignoring flanking letters. Each trial began with a fixation cross in the center of the screen for 500 ms, followed by the flanker letters for 400 ms. The target letter was then presented in the center of the screen along with the flanker letters until the participant pressed one of two keys corresponding to either possible target letter, or until the trial timed out if no response was detected within 6000 ms. Each trial was followed by a blank inter-trial period, which varied in duration between two different block types: "behavioural" blocks (250 ms) and "pupil dilation" blocks (2000 ms).

The former block type allowed data relating to task performance to be collected without unnecessarily introducing fatigue due to longer inter-trial latency, whereas the latter block type necessarily introduced this latency and was more specifically catered to pupillometric data collection. Participants completed 16 initial practice trials of the behavioural block, before completing 152 trials of each block type, which were divided into four 76 trial blocks. Within each block, combinations of the target letter and flanker letters were congruent (e.g. AAAAA) for half of the trials, and incongruent (e.g. AALAA) for the other half. The two possible target letters varied between blocks, with the Behavioural block type having one set of target letters and the Pupillometry block type having another. Within each block type, each possible target letter occurred an equal number of times (see **Figure 2**).

Psychological questionnaire measures

***Narcissism*.** Both grandiose and vulnerable narcissism were assessed using the 60-item Five-Factor Narcissism Inventory Short Form (Sherman et al., 2015). Participants responded to each statement on a 5-point Likert scale ranging from 1 (disagree strongly) to 5 (agree strongly). The questionnaire consists of two primary sub-scales containing items assessing grandiose (e.g. "Others say I brag too much, but everything I say is true") and vulnerable narcissism (e.g. "I often feel as if I need compliments from others to be sure of myself") respectively.

***Aggression*.** Trait aggression was assessed using the 29-item Buss-Perry Aggression Questionnaire (Buss & Perry, 1992). Participants responded to each statement on a 5-point scale ranging from 1 (extremely uncharacteristic of me) to 5 (extremely characteristic of me). The questionnaire consists of four sub-scales, two of which were used in the present study; anger (e.g. "I flare up quickly but get over it quickly") and hostility (e.g. "I am suspicious of overly friendly strangers").

***Borderline personality disorder*.** Borderline personality disorder traits were assessed using the 10-item McLean Screening Instrument for Borderline Personality Disorder (Zanarini et al., 2003). Participants responded to each item with either "Yes" or "No". Each item assesses a different diagnostic trait of borderline personality disorder (e.g. "Have you been extremely moody?", "Have you chronically felt empty?"). A delay with obtaining ethical approval for this scale's inclusion meant that the McLean Screening Instrument for Borderline Personality Disorder questionnaire could only be completed by 61 out of the 100 participants.

Data processing and analysis

Pupil dilation data was cleaned using a band pass filter to remove extreme artifacts, followed by linear interpolation to remove gaps in the data caused by blinks and brief signal loss. A total of 95 trials (0.2% of total trials) were excluded as they contained too little data to correctly perform interpolation, as were trials where pupil dilation data was irrelevant (i.e. feedback trials). Finally, all pupil dilation data was converted from arbitrary units of area to z-scores within-trials, in order to standardize the data and control for any effect of baseline pupil dilation on the observed change in amplitude. In order to process the data for analysis, data within each trial was cut down to a target window that began 50 milliseconds prior to the stimulus and continued for the rest of the duration of the trial. Within this target window, baseline pupil dilation for each trial was defined as the pupil diameter at the beginning of the target window, and peak pupil dilation was defined as the greatest pupil diameter within 400 milliseconds of the beginning of the target window (Holmqvist et al. 2011). Change in amplitude between baseline and peak was calculated using these values. For each participant, pupil dilation data for all trials was then averaged within-blocks. Data relating to accuracy for both block types (i.e. number of correct / incorrect responses) were converted into D-prime (D') statistics to correct for response bias. Some participants (N = 18) had a 100% hit-rate and/or a 0% false alarm rate. As it is not possible to calculate d' with values of 1 or 0, these values were adjusted to the effective limit of 0.99 or 0.01 respectively. Finally, to maximize trait

differences for between-group comparisons, participants were divided into three groups depending on whether they were in the lower, middle or top third of scorers on measures of overall trait narcissism, vulnerable narcissism and grandiose narcissism.

In order to analyse the independent within-subjects effect of flanker task trial type, three one-way ANOVA analyses were conducted where the independent variable was flanker task stimuli congruency (congruent, incongruent), and the dependent variable for each analysis was either pupil dilation magnitude, D' or reaction time respectively.

Similarly, to analyse the independent effects of narcissistic trait level, a series of nine one-way ANOVA analyses was conducted where the independent variable for each analysis was either overall narcissistic trait level (low, medium, high), grandiose narcissistic trait level (low, medium, high), or vulnerable narcissistic trait level (low, medium, high) respectively. For each independent variable, the analysis was repeated for each dependent variable, which was either pupil dilation magnitude, D' or reaction time respectively.

Finally, to analyse interaction effects between flanker task trial type and narcissistic trait level, a series of nine two-way mixed 2x3 ANOVA analyses was conducted, where the within-subjects variable was flanker task stimuli congruency (congruent, incongruent), and the between-subjects variable for each analysis was either overall narcissistic trait level (low, medium, high), grandiose narcissistic trait level (low, medium, high), or vulnerable narcissistic trait level (low, medium, high) respectively. For each combination, the analysis was repeated for each dependent variable, which was either pupil dilation magnitude, D' or reaction time respectively.

4.2.ii. Results

Table 1. Descriptive statistics for narcissistic trait groups and psychometric measures of aggression.

		M	*SD*	*Minimum*	*Maximum*	*α*
FFNI Narcissism Total	High	176.70	14.95	162	223	.852
	Medium	153.20	5.44	145	161	
	Low	132.57	9.19	100	143	
FFNI Vulnerable Narcissism	High	58.55	4.99	52	74	.681
	Medium	47.71	2.02	45	51	
	Low	38.97	4.08	31	44	
FFNI Grandiose Narcissism	High	124.67	14.6	110	174	.884
	Medium	102.58	3.60	96	109	
	Low	84.26	7.81	67	95	
BPAQ (Total)	-	68.76	13.98	42	108	.851
MSI BPD	-	3.58	2.48	0	9	.703

α = Cronbach's alpha

Between-groups descriptives and correlations

Descriptive statistics for all narcissistic trait groups and psychometric measures of aggression are presented in **Table 1**. A matrix of zero-order correlations between narcissistic traits, trait aggression and BPD traits are presented in **Table 2**. A significant positive correlation was observed between overall trait narcissism, overall

trait aggression and BPD traits. Vulnerable narcissism was positively correlated with trait aggression and BPD traits, whereas grandiose narcissism was positively correlated only with trait aggression. BPD traits were positively correlated with overall aggression.

Table 2. Zero order correlations between narcissistic traits and psychometric measures of aggression.

	FFNI VN	FFNI GN	BAQ Total	MSI BPD
Five-Factor Narcissism Inventory (Total)	-	-	.510**	.320*
Five-Factor Narcissism Inventory (Vulnerable)	-	-.014	.576**	.583**
Five-Factor Narcissism Inventory (Grandiose)	-	-	.340**	.170
Buss-Perry Aggression Questionnaire (Total)	-	-	-	.660**

* $p < 0.05$, ** $p < 0.01$

A matrix of zero-order correlations between all narcissistic traits, trait measures of aggression, and dependent variables are presented in **Table 3**. No significant correlations were observed between either overall trait narcissism, overall trait aggression or BPD traits, and change in pupil dilation magnitude, D' or reaction time.

Within-groups descriptives and correlations

Across all trials, pupil dilation magnitude was low but positive (i.e. indicating an increase in pupil dilation; *M* = .73, *SD* = .28) and varied little, with the mean magnitude for incongruent trials (*M* = .74, *SD* = .27) being only marginally greater than the magnitude for congruent trials (*M* = .73, *SD* = .28). Reaction time (ms) also varied little (*M* = 398.20, *SD* = 39.94), with incongruent trials (*M* = 401.09, *SD* =

57.46) having only marginally longer reaction times than congruent trials (*M* = 395.30, *SD* = 48.99). D' scores were very high (i.e. indicating very high task accuracy) and similarly non-variable (*M* = 4.09, *SD* = .48), with little difference between incongruent (*M* = 4.07, *SD* = .48) and congruent (*M* = 4.11, *SD* = .51) trials. Task performance scores were more variable, with scores across congruent trials being considerably worse compared to incongruent trials. A significant positive correlation was observed between D' scores and reaction time across all trials (*r* = .235, *p* = .026).

Table 3. Zero order correlations between narcissistic traits, trait measures of aggression, and pupil dilation magnitude, D' and reaction time.

	PD	D'	RT
Five-Factor Narcissism Inventory (Total)	.137	-.012	.134
Five-Factor Narcissism Inventory (Vulnerable)	.148	-.053	.062
Five-Factor Narcissism Inventory (Grandiose)	.128	.031	.108
Buss-Perry Aggression Questionnaire (Total)	.189	-.024	.058
MSI Borderline Personality	.025	.055	-.001

Comparison of within-subjects variables

No effect of flanker task stimuli congruency was observed for pupil dilation magnitude ($F(1, 98) = .66$, $p = .41$), D' scores ($F(1, 99) = .38$, $p = .53$) or reaction time ($F(1, 85) = .102$, $p = .75$).

Comparison of between-subjects variables

No significant main effects were observed for overall trait narcissism, grandiose narcissism or vulnerable narcissism (see **Appendix A1**). No significant interaction

effects were observed between flanker stimuli congruency and the three types of narcissistic trait level (see **Appendix A1**).

4.2.iii. Interim Discussion

The first in the series of two experiments described by the present study provided largely inconclusive results, with no task-related to trait-related effects observed. Given the strong and theoretically coherent relationships observed between trait measures, it is unlikely that the lack of trait-related effects is due to a failure of the psychometric measures featured in the experiment. Rather, the lack of trait-related effects is more likely a secondary consequence of the lack of task-related effects, which may indicate methodological issues with the flanker task. This is not to suggest that the lack of hypothesized task-related effects necessarily indicates methodological failure, but rather that there are several fundamental task-related effects that should occur as a feature of the task independently of hypotheses. Specifically, there was no effect of stimuli congruency observed, suggesting the interference effect assumed to be produced by the flanker task was not present. As this effect would be necessary to elicit conflict detection, its absence would naturally lead to the lack of task-related and trait-related effects. An examination of the data suggests this may be attributable to participants not finding the task sufficiently challenging, with D' scores observed to be very high, and reaction times being faster than expected (McDermott et al., 2017; Wühr & Heuer, 2017). If participants did not struggle to ignore flanking stimuli and correctly identify target stimuli, then any manipulations of stimuli congruency would be ineffectual. The first experiment was therefore unfortunately unable to confirm, reject or develop the hypotheses it was intended to address (see **General Discussion** for further reflection).

4.3. Experiment 2

4.3.i. Method

Participants

Participants were 92 students (16 male, 76 female) aged between 18 and 31 years (M = 21.59, SD = 3.08) attending the University of Nottingham. Participants were either Psychology students taking part in order to earn research credits, or students recruited from advertisements distributed via email. Participants were told they would be taking part in an eye-tracking study investigating the link between cognition, physiology and behaviour. The experiment was approved by the School of Psychology ethics committee (Ref: S1195).

Procedure

Participants were asked to fill out the Five Factor Narcissism Inventory (Short Form; Sherman et al. 2015), the Buss-Perry Aggression Questionnaire (Buss & Perry 1992) and the McLean Screening Instrument for Borderline Personality Disorder (Zanarini et al. 2003), at either the start or the end of the study. Participants then completed the experimental task using a computer running Experiment Builder (version 2.1.140; SR Research 2017).

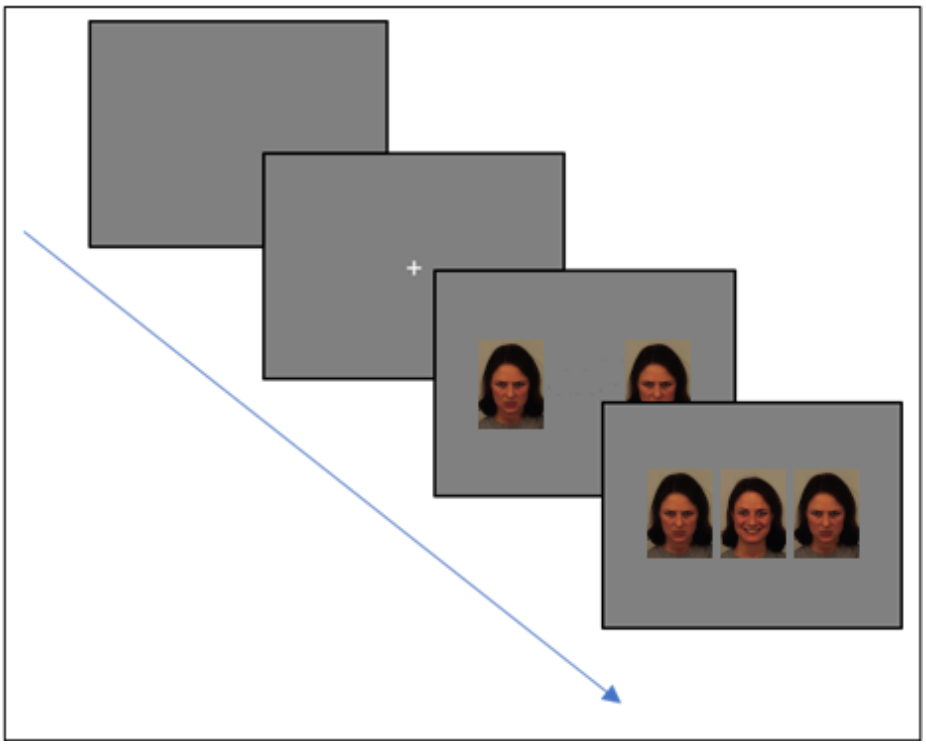
Eye-tracking specifications

Pupil diameter was measured throughout the experiment using EyeLink 1000 (SR Research 2017) eye-tracking hardware. The specifications used in the present study were the same as those used in the previous study (see **4.4.1.i**).

Behavioural task

Emotion face flanker task. Participants were required to respond correctly to one of three possible target faces by pressing a corresponding letter on a keyboard, whilst ignoring flanking faces and face gender/identity. Each face was either male or female and had an expression that was categorized as either angry, neutral or happy. Faces were adapted from the Karolinska Directed Emotional Faces database (Calvo & Lundqvist, 2008). Each trial began with a fixation cross in the center of the screen for 500 ms, followed by the flanking faces for 400 ms.

Figure 3. Diagram showing the phases of an incongruent trial on the face-based flanker task showing angry emotion faces as flanking stimuli paired with a happy emotion face target stimuli.



The target face was then presented in the center of the screen along with the flanking faces until the participant pressed one of the three possible keys, or until the trial timed out if no response was detected within 6000 ms. Each trial was followed by a

blank inter-trial period, which varied in duration between the two "behavioural" (250 ms) and "pupil size" (2000 ms) block types as described in previous experiment. Participants completed 16 initial practice trials, before completing 144 trials of each block type, which were divided into four 72 trial blocks. Within each block, combinations of the target and flanking face emotions were congruent (e.g. angry - angry - angry) for half of the trials, and incongruent (e.g. angry - happy - angry) for the other half. The identity of the face stimulus was consistent within each trial (i.e. the identity of the flanking and target faces were consistent). The gender and identity of each face stimulus was randomized. Within each block type, each possible type of target face emotion occurred an equal number of times (see **Figure 3**).

Data processing and analysis

Pupil dilation data was cleaned using a band pass filter to remove extreme artifacts, followed by linear interpolation to remove gaps in the data caused by blinks and brief signal loss. Feedback trials were excluded. One participant was excluded as the eye-tracker struggled to reliably track their pupil, resulting in too little usable data to correctly perform interpolation. Finally, all pupil dilation data was converted from arbitrary units of area to z-scores within-trials, in order to standardize the data and control for any effect of baseline pupil dilation on the observed change in amplitude. In order to process the data for analysis, data within each trial was cut down to a target window that began 50 milliseconds prior to the stimulus and continued for the rest of the duration of the trial. Within this target window, baseline pupil dilation for each trial was defined as the pupil diameter at the beginning of the target window, and peak pupil dilation was defined as the greatest pupil diameter within 400 milliseconds of the beginning of the target window. Change in amplitude between baseline and peak was calculated using these values. For each participant, pupil dilation data for all trials was then averaged within-blocks (i.e. for each type of facial emotion stimulus). As calculating D' was not possible for this dataset given the number of different target/flanker permutations per trial (i.e. stimuli were not merely congruent or incongruent, but also featured different types of incongruency), data

relating to accuracy were instead converted into Z-scores and transformed to a basic signal-to-noise ratio (i.e. number of correct / incorrect responses) to correct for response bias (this data is henceforth referred to as 'task performance'). Finally, participants were divided into three groups for between-group comparisons depending on whether they were in the lower, middle or top third of scorers on measures of overall trait narcissism, vulnerable narcissism and grandiose narcissism.

In order to analyse the independent within-subjects effects of task variables, a series of nine one-way ANOVA analyses was conducted, where for each of the analyses the independent variable was either flanker stimuli congruency (congruent, incongruent), flanker face type (angry, happy, neutral) or target face type (angry, happy, neutral) respectively. For each independent variable, the analysis as repeated for each dependent variable, which was either pupil dilation magnitude, task performance or reaction time respectively.

Similarly, to analyse the independent effects of narcissistic trait level, a series of nine one-way ANOVA analyses was conducted where the independent variable for each analysis was either overall narcissistic trait level (low, medium, high), grandiose narcissistic trait level (low, medium, high), or vulnerable narcissistic trait level (low, medium, high) respectively. For each independent variable, the analysis as repeated for each dependent variable, which was either pupil dilation magnitude, task performance or reaction time respectively.

Finally, to analyse interaction effects between within-subjects variables and narcissistic trait level, an extensive series of two-way mixed ANOVA analyses was conducted. For the first set of analysis in this series, nine two-way mixed 2x3 ANOVA analyses were conducted, where the within-subjects variable was flanker task stimuli congruency (congruent, incongruent), and the between-subjects variable for each analysis was either overall narcissistic trait level (low, medium, high), grandiose narcissistic trait level (low, medium, high), or vulnerable narcissistic trait level (low, medium, high) respectively. For each combination, the analysis was

repeated for each dependent variable, which was either pupil dilation magnitude, task performance or reaction time respectively. For the remaining analyses, a series of two-way mixed 3x3 ANOVA analyses was conducted, where the within-subjects variable was either flanker face type (angry, happy, neutral) or target face type (angry, happy, neutral) respectively, and the between-subjects variable for each analysis was either overall narcissistic trait level (low, medium, high), grandiose narcissistic trait level (low, medium, high), or vulnerable narcissistic trait level (low, medium, high) respectively. For each combination, the analysis was repeated for each dependent variable, which was either pupil dilation magnitude, task performance or reaction time respectively.

4.3.ii. Results

Table 4. Descriptive statistics for narcissistic trait groups and psychometric measures of aggression.

		M	*SD*	*Minimum*	*Maximum*	*α*
FFNI Narcissism Total	High	190.57	10.27	177	215	.882
	Medium	162.07	6.91	152	175	
	Low	138.41	10.31	117	151	
FFNI Vulnerable Narcissism	High	60.23	4.22	55	73	.699
	Medium	50.53	2.36	47	54	
	Low	40.79	3.89	32	46	
FFNI Grandiose Narcissism	High	132.06	9.66	121	156	.888
	Medium	107.28	7.88	96	119	
	Low	85.34	8/69	61	95	
BPAQ (Total)	-	72.38	14.65	40	108	.846
MSI BPD	-	3.11	2.47	0	10	.729

α = Cronbach's alpha

Between-subjects descriptives and correlations

Descriptive statistics for all narcissistic trait groups and psychometric measures of aggression are presented in **Table 4**. A matrix of zero-order correlations between narcissistic traits, trait aggression and BPD traits are presented in **Table 5**. A significant positive correlation was observed between overall trait narcissism and overall trait aggression. Vulnerable narcissism more specifically was positively associated with overall trait aggression and BPD traits, whereas grandiose narcissism was positively associated only with overall trait aggression. Finally, overall trait aggression was positively correlated with BPD traits.

Table 5. Zero order correlations between narcissistic trait groups and psychometric measures of aggression.

	FFNI VN	FFNI GN	BAQ Total	BPD
Five-Factor Narcissism Inventory (Total)	-	-	.521**	.098
Five-Factor Narcissism Inventory (Vulnerable)	-	.078	.430**	.309**
Five-Factor Narcissism Inventory (Grandiose)	-	-	.446**	.046
Buss-Perry Aggression Questionnaire (Total)	-	-	-	.331**

** $p < 0.01$

A matrix of zero-order correlations between all narcissistic traits, trait measures of aggression, and dependent variables are presented in **Table 6**. A significant negative correlation was observed between overall trait narcissism and pupil dilation magnitude, as well as between vulnerable narcissism and pupil dilation magnitude, and between trait aggression and pupil dilation magnitude.

Table 6. Zero order correlations between narcissistic traits, trait measures of aggression, and pupil dilation magnitude, task performance, and reaction time.

	PD	ACC	RT
Five-Factor Narcissism Inventory (Total)	-.238*	-.102	.108
Five-Factor Narcissism Inventory (Vulnerable)	-.288**	-.052	-.087
Five-Factor Narcissism Inventory (Grandiose)	-.167	-.098	.150
Buss-Perry Aggression Questionnaire (Total)	-.247*	-.042	.030
MSI Borderline Personality	-.146	.170	-.075

* $p < 0.05$, ** $p < 0.01$

Within-subjects descriptives and correlations

Descriptive statistics for all within-subjects variables are presented in **Table 7.** Across all trials, pupil dilation magnitude was positive (i.e. indicating an increase in pupil dilation) and relatively non-variable. Reaction time (ms) was similarly consistent across all trials. Task performance scores were more variable, with scores across congruent trials being considerably worse compared to incongruent trials. Task performance across all trials was relatively poor, with scores for some trial types being close to chance. A significant positive correlation was observed between task performance scores and reaction time across all trials (r = .256, p = .015). A significant negative correlation was observed between pupil dilation amplitude and task performance scores across all trials (r = -.311, p = .003).

Comparison of within-subjects variables

A significant main effect of flanker stimuli congruency on task performance scores was observed ($F(1, 88) = 7.30$, p = .008), with scores being higher for incongruent

trials. No effect was observed for pupil dilation magnitude ($F(1, 89) = .89$, $p = .34$) or reaction time ($F(1, 87) = 1.30$, $p = .25$).

A significant main effect of flanker face type on reaction time was observed ($F(1, 89) = 16.53$, $p<.001$). Pairwise comparisons indicated reaction time was significantly lower when the flanker face type was 'happy compared to both when the face type was 'neutral' ($p<.001$) or 'angry ($p<.001$). No effect of flanker face type on pupil dilation magnitude ($F(2, 178) = 2.33$, $p = .10$), task performance scores ($F(2, 154) = .42$, $p = .65$) was observed.

Table 7. Descriptive statistics for pupil dilation magnitude, task performance, and reaction time

	PD		Task Performance		RT (ms)	
	M	*SD*	*M*	*SD*	*M*	*SD*
All Trials	2.95	.53	2.02	1.36	879.37	201.86
Congruent Trials	2.97	.53	1.58	1.70	882.29	212.45
Incongruent Trials	2.96	.53	2.68	2.55	872.42	193.31
Angry - Flanker	3.00	.55	1.84	2.04	888.54	233.72
Happy - Flanker	2.96	.56	1.75	1.88	839.99	191.06
Neutral - Flanker	2.92	.54	1.68	1.64	897.63	194.78
Angry - Target	3.03	.55	1.73	1.79	875.70	198.40
Happy - Target	2.93	.52	1.66	1.62	881.93	216.75
Neutral - Target	2.92	.54	1.65	1.69	869.67	190.32

A significant main effect of target face type on pupil dilation magnitude was observed ($F(2, 178) = 11.73$, $p<.001$). Pairwise comparisons indicated that pupil dilation magnitude was significantly higher when the target face type was 'angry' compared to both when the face type was 'happy' ($p<.001$) or 'neutral' ($p<.001$). A significant main effect of target face type on reaction time was observed ($F(2, 174) = 3.40$, $p = .036$). Pairwise comparisons indicated that reaction time was significantly lower when the target face type was 'neutral' compared to both when the face type was 'happy' ($p = .034$). No effect was observed for task performance scores ($F(2, 150) = .048$, $p = .953$).

Comparison of between-subjects variables

No significant main effects were observed for overall trait narcissism, vulnerable narcissism or grandiose narcissism (see **Appendix A2**). No significant interaction effects were observed between any of the three types of narcissistic trait level and either trait aggression level and/or borderline personality trait level (see **Appendix A2**). Similarly, no significant interaction effects were observed between the three types of narcissistic trait level, trait aggression level, and/or borderline personality trait level and either flanker face type, target face type, target face type when the flanker face type was 'angry', or flanker stimuli congruency (see **Appendix A2**).

4.3.iii. Interim Discussion

Where the first experiment produced largely inconclusive results, the second experiment proved more successful. This may be attributable to the increased difficulty of the task, which allowed interference effects to become more prominent. Given the considerably greater magnitude of changes to pupil dilation observed, corresponding with considerably lower task performance scores and greater reaction time, this suggestion seems plausible, and more importantly affirms the hypothesized relationship between pupil dilation and conflict detection. The task-relevant effects observed further affirm this relationship, and lend credence to the hypothesis that

conflict detection may underlie threat sensitivity through attributing more salience to conflicts involving potentially threatening stimuli. However, methodological issues remain a concern, with the very low performance scores, unusual effect of congruence, and lack of trait-related effects presenting points for consideration (see **General Discussion**).

4.4. General Discussion

Evaluation of trait-independent effects on task performance

No task-related effects on behaviour were observed in the first experiment, with flanker task stimuli congruency having no apparent influence on pupil dilation magnitude, reaction time or task accuracy (D' score). Given the relatively less established use of pupil dilation magnitude as a measure for assessing conflict detection in a flanker task, the lack of task-related effects on this measure is not necessarily unusual. However, the lack of task-related effects on task accuracy and reaction time is more unusual, given the straightforward relationship between these measures and conflict-related interference.

A simple potential explanation for this lack of an effect of stimuli congruency is that participants were able to focus on target stimuli without being distracted by flanking stimuli. If this was the case, then it would be expected that the content of flanking stimuli would not be of any circumstance, and therefore manipulating congruency would not have any effect on task performance. However, as an assumption of the task is that incongruence between flanking and target stimuli should create a conflict that is distracting and has some influence on task performance, a lack of sufficiently distracting incongruence would indicate that the task did not function as intended. A descriptive examination of the data lends credibility to this explanation, with participants in the first experiment exhibiting very high D' scores across congruent and incongruent trials and across groups. This suggests participants had very little difficulty making a correct response, and as a result it would stand to reason that

interference effects caused by stimuli incongruency (if any) would be very weak. It is possible that more complex task permutations would have been more challenging for participants and therefore more successful at eliciting conflict-related interference effects. For example, altering stimulus-response compatibility (e.g. pressing 'A' when target stimuli is 'L', and vice versa) in order to introduce an additional layer of response inhibition, or using directional stimuli instead of letter-based stimuli which may facilitate more pronounced interference effects when combined with alterations to stimulus-response compatibility (Ridderinkhof et al., 2021).

It is of course not the case that being more difficult arbitrarily improves the task a measure of conflict detection, as where the letter-based flanker task was too easy, another task may have been too difficult, which would similarly result in task manipulations (i.e. congruence) having little effect. However, the presence of task-related effects observed in the second experiment suggests that task difficulty (or a lack thereof) was a meaningful factor in the present study, that may have contributed to the relative success of the second experiment to yield more conclusive results. Specifically, the second experiment involved a flanker task featuring face emotion stimuli. In this task, participants had to choose between one of three possible responses instead of two, and identify stimuli without the clearly discernible features of letter-based stimuli. As a result, the task was necessarily more difficult, and this is reflected in the much lower task performance scores observed for participants in the second experiment across all trials and across groups.

The task-related effects observed relate to two different types of conflict-related effect potentially present in the face-based flanker task used in the present study. The first is a standard flanker interference effect caused by stimuli congruency, whereas the second is an effect of emotional valence that may amplify the salience of a conflict and influence the magnitude of any interference effects. In accordance with the latter, it was observed that the emotional valence of stimuli was influential, with trials featuring angry faces as target stimuli being associated with significantly higher pupil dilation magnitude, and trials featuring happy faces as flanking stimuli being

associated with significantly lower reaction time. Happy faces may be perceived as unambiguously non-threatening, and may therefore produce a weaker interference effect by virtue of being less distracting. Interestingly, happy emotional faces as target stimuli were also associated with a significantly lower reaction time compared to neutral faces, which may be an effect of ambiguity as discussed in the introduction. That is to say, this may represent that neutral faces were in general harder to distinguish from angry faces due to perceptual biases, whereas this perceptual bias did not apply to happy faces, resulting in happy emotional faces being easier to identify. The association between trials featuring angry faces as target stimuli and higher pupil dilation magnitude may be attributable to angry faces being perceived as more salient due to signalling implicit indications of threat. As a result, detected conflicts involving angry faces may trigger a similarly salient conflict detection signal (Horstmann & Bauland, 2006). If pupil dilation is taken to be representative of sympathetic arousal linked to conflict detection, then increased pupil dilation magnitude in response to trials featuring angry faces is consistent with the assumption that angry faces should attract more attention irrespective of individual differences (Huang et al., 2011; Pérez-Dueñas et al., 2014). However, if this description of the effect is correct, then it would be expected that trials featuring angry faces as target stimuli should be associated with enhanced performance or faster reaction time, whereas trials featuring angry faces as incongruent flanking stimuli should be associated with worse performance or slower reaction time (Horstmann et al., 2006). In contrast, no task-related effects of this kind were observed. Furthermore, it would be expected that increased pupil dilation magnitude would be similarly observed for trials featuring angry faces as flanking stimuli, whereas this effect was not observed. There is perhaps some evidence of this effect, in that pupil dilation magnitude was found to be negatively correlated with task performance, which could be taken to represent a link between pupil dilation and conflict-related interference (i.e greater arousal in response to salient emotional conflicts causes these conflicts to become more distracting, resulting in worse overall performance).

It may therefore be the case that changes in pupil dilation represent a physiological response associated with conflict detection, and that this response is sensitive to the salience of the conflict detected, but that this effect only loosely manifested in task performance in the present study. This is plausible when one considers that changes in pupil dilation and changes in task performance are two related but nevertheless distinct responses to conflict detection. Whilst pupil dilation is more insulated from situational factors by virtue of being an autonomic physiological response, changes to task performance are a product not only of cognitive conflict-related processes (e.g. interference) but are also a product of practical features of the task. For example, if the task featured in the second experiment was indeed too difficult, which is certainly arguable given the low task performance scores, then any task-related effects on performance (e.g. the emotional content of stimuli) would be confounded by this high difficulty. In contrast, task-related effects on pupil dilation magnitude would be relatively unaffected by the difficulty of the task, as they are influenced solely by the presence and salience of a conflict. Interestingly, this separation highlights the nuanced nature conflict-related changes to pupil-dilation with relation to the face-based flanker task, wherein enhanced sensitivity to a conflict might reflect either enhanced or impaired performance depending on the required task-related behaviour that follows (i.e. an angry face may be easier to identify as a target, but it may also be more distracting as a flanker).

The suggestion that the task featured in the second experiment was too difficult is supported by the unexpected observation that task performance in the face-based flanker task was better on incongruent trials than congruent trials, which naturally runs counter to the assumption that incongruency should create an interference effect that impairs performance. This may indicate that the facial emotion images used as stimuli were too difficult to correctly identify, to the extent that it actually became easier to identify them when supported by a contrasting incongruent flanking stimuli. For example, it may have been difficult to correctly identify whether a facial emotion was angry or neutral, unless compared with an incongruent facial emotion that was more distinctly angry or neutral, providing a useful reference for making a more

accurate decision. Whilst the face images used as the stimuli in the task were designed to be distinct from one another, it may be that the variability of different facial identities acted as a confound making it more challenging for participants to calibrate their assessments of emotion type. An alternative to the Karolinska Directed Emotional Faces database is an averaged database produced from the same face images, smoothed to reduce the influence of identity and create a cleaner presentation of facial emotion. It is possible that using averaged face images such as these for stimuli would have made the task sufficiently challenging to elicit task-related effects. It is interesting to note that task performance scores were also relatively variable, suggesting that participant experience of task difficulty was not uniform. This may be attributable to unclear task instructions or insufficient practice trials failing to successfully prepare all participants to the same extent. It may also be attributable to an uncontrolled and unidentified variable that influenced task performance. Given the long duration of the flanker task, it could be that fatigue (and participant tolerance to fatigue) became an influential factor that may have interacted with and amplified task difficulty for some participants. None of the above issues were flagged in feedback provided by participants involved in a limited pilot test of the flanker tasks involved in either the first or second experiment, but it is nevertheless possible that there may have been methodological issues with both tasks that became evident in a larger sample.

Evaluation of individual differences and trait-related effects

No trait-related effects on behaviour were observed in either the first or second experiment. Psychometric measures relating to narcissistic aggression were included to verify that hypothesized associations between traits were present. The results from both experiments do not suggest that the lack of an effect is attributable to failure of the psychometric measures used in either experiment, as the correlation between traits observed were consistent both with theoretical assumptions and the previous studies. This also suggests no necessary concern for the effect of multiple comparisons. Specifically, no correlation was observed between vulnerable and grandiose

narcissism, vulnerable narcissism was uniquely correlated with BPD traits, and all traits were associated with overall trait aggression. This suggests that participants were accurately described by trait measures, and therefore it is not the case that trait-related effects were incorrectly characterized, but rather effects were not strongly trait-related (whether due to issues with the task or genuine lack of effects). The term 'strongly' is chosen deliberately, as a correlation was observed between pupil dilation magnitude and overall trait narcissism, vulnerable narcissism and trait aggression, with these traits being associated with significantly lower pupil dilation magnitude across trials. If this correlation were positive rather than negative, then this relationship would be consistent both with the hypotheses regarding the influence of these traits on conflict detection, and the results discussed above suggesting pupil dilation magnitude may be associated with conflict detection. However, as this is not the case, this observed correlation is unexpected and apparently runs counter to the hypotheses. If pupil dilation magnitude is taken to represent sympathetic arousal associated with detection of a conflict, then this finding would suggest that these trait measures were associated with reduced conflict-related arousal. This effect is difficult to interpret, especially in the light of any related between-groups trait-related effects that would lend confidence to this effect. It could be that this negative relationship reflects an association between antisocial and aggressive traits and low physiological arousal at rest that has been documented in the literature (Lorber, 2004; Raine, 2002). This might not intuitively be expected to apply, as the measure of pupil dilation featured in the present study represents a measure of active physiological arousal in a task context designed to elicit cognition and behaviour relevant to narcissistic aggression. However, if the task featured in the second experiment was for some reason not appropriate for eliciting the hypothesized trait-related effects of narcissistic aggression, then it is plausible to suggest a residual effect of this baseline tendency towards lower physiological arousal may have caused a negative relationship to be observed between relevant trait measures and pupil dilation magnitude. Whilst the decision to use facial emotion stimuli to investigate narcissistic aggression in a more 'socially-relevant' task context (i.e. compared to the more straightforward stimuli featured in the first experiment) was theoretically robust (De

Panfilis et al., 2019), this approach was nevertheless exploratory, and may not have been sufficiently socially-relevant to facilitate cognition and behaviour relevant narcissistic aggression. In contrast, the comparably more overt approach of Sleegers et al. (2017) using ostracism could have been more effective. This would suggest that it may not be possible to wholly reduce narcissistic threat sensitivity to conflict detection beyond a certain context-sensitive threshold, even when the task context in which one is examining conflict detection has features relevant to narcissistic cognition.

Evaluation of pupillometry as an operationalisation of narcissistic aggression

The use of pupillometry in the present study was a highly exploratory method intended to be a metric of conflict detection, due to the link between pupil dilation and changes in cognitive control. The magnitude of pupil dilation was hypothesized to be a physiological response representing the extent to which cognitive control was increased following the detection of a conflict, and therefore act as a measure of conflict detection (i.e. greater sensitivity to conflicts should be associated with greater increases in cognitive control). Pupil dilation is a challenging measure to work with, being a dynamic physiological response that is influenced by various different factors (Sirois & Brisson, 2014), from abstract influences such as emotion, to cognitive processes (e.g. cognitive control), to basic environmental factors such as light level. Luminance was controlled as much as possible in both experiments. The more diverse appearance of stimuli in the second experiment (i.e. face stimuli) may have resulted in changes in luminance being less controlled between trials, however it seems likely any changes would have been very minor, given the appearance of stimuli remained largely standardized. The dynamic nature of pupillometric data necessitates a careful and deliberate approach to cleaning and formatting the data for analysis. The approach taken for both the experiments documented in the present study involved the extraction of a finite target window for each trial, within which changes in pupil dilation magnitude were calculated and recorded. Whilst the size of this window was defined based on guidance from the literature (Holmqvist et al., 2011; Laeng et al.,

2012), pupillometry is not in every sense an exact science, and it is possible that this window was too conservative. For example, other researchers have observed relevant changes in pupil dilation occurring with a relatively long latency, which may feasibly have been overlooked in the approach taken in the present study (Hubert, Lyall and Järvikivi, 2021). However, given the exploratory nature of the study, adopting a more conservative approach was ultimately considered most appropriate, in order to be allow enough scope for relevant changes in pupil dilation to be observed whilst remaining confident that any effects observed were indeed related to the task.

Despite the challenging nature of pupillometry, the second experiment documented a task-related effect on pupil dilation magnitude, in accordance with hypotheses concerning the relationship between pupil dilation, conflict detection, and the salience of conflict-related content, and coherent with previous research reporting similar findings (Hershman et al., 2021; Laeng et al., 2011; van der Wel & van Steenbergen, 2018; Van Steenbergen & Band, 2013). In the first experiment, pupil dilation magnitude was observed to be largely non-variable, suggesting that if there were any relevant trial-related changes these were very minor, and indeed no task-related or trait-related effects were observed in the first experiment. In the second experiment, pupil dilation magnitude was much greater across trials, and a task-related effect was observed as described above, however trait-related effects were not observed. In addition to the suggested explanation for this lack of an effect provided above in terms of task design, it may also be the case that pupillometry is not the most appropriate measure for studying conflict detection in narcissistic aggression. The typically small effect sizes associated with individual differences effects (Gignac & Szodorai, 2016) may render pupillometric effects difficult to capture, with similar oculometric research studying individual differences reporting small effect sizes (Unsworth et al., 2019). However, it must also be acknowledged that other research similar to the present study has observed clear effects irrespective of this (Al-Samarraie et al., 2018; Hubert Lyall & Järvikivi, 2021; Inzlicht et al., 2015; Sleegers et al., 2017; Unsworth et al., 2019). On the basis of this prior research and given the relatively strong effects of trait narcissism observed in the previous studies

documented in the thesis, it was not assumed that individual differences-related effects on pupil dilation would be especially challenging to capture. Furthermore, regardless of this, the large sample size featured in both experiments was a means of pre-emptively counteracting any obfuscation due to small effect sizes. Nevertheless, this potentially weak effect size of pupillometric effects in the context of individual differences may have contributed to the lack of trait-related effects observed.

Conclusion

Whilst issues with the behavioural task used in the first experiment described in the present study limited the ability to observe task-related effects relevant to the hypothesised association between pupil dilation and conflict detection, the second experiment was more conclusive. Identifying images of angry faces (i.e. potentially threatening emotional stimuli) was associated with pupil dilation responses of a greater magnitude, and pupil dilation magnitude was more generally correlated with worse task performance. However, this effect was not observed when examining the more fundamental effect of stimuli congruency, and complimentary corresponding effects of stimuli type on task performance or reaction time were largely absent. The task-related effect observed in the second experiment is therefore ultimately inconclusive, but provides encouraging evidence to support the hypothesis that pupil dilation functions as a measure of conflict detection that is sensitive to potentially threatening (or salient) stimuli.

The largely absent trait-related effects observed in either of the experiments described in the present study may be attributable to a combination of apparent methodological issues affecting the flanker task, practical issues with the use of pupillometry to study trait-related effects in a situation that is not heavily influenced by those traits, and a lack of the hypothesized effects between narcissistic aggression and conflict sensitivity. Whilst theoretically justified, narcissistic threat sensitivity may not be reducible to basic conflict sensitivity outside of more explicitly threatening and explicitly social contexts (e.g. involving ostracism or insult). This would account for

the lack of an observed effect in the first experiment considering the nature of the stimuli used, but it is interesting that an effect was not apparent in the second experiment featuring stimuli that one would expect to be relevant to interpreting potential social threats (i.e. spotting facial emotion signals that violate expectations and responding accordingly should be a trait feature of narcissism that facilitates narcissistic aggression). This may be due to the relatively weak effect of trait influences on pupil dilation, and the need to examine these influences in the context of a task that is not merely relevant but strongly and explicitly related to narcissistic aggression. The attempt to reduce measures of narcissistic aggression to simplified and more objective methods that constitutes the broader aim of the present study was therefore met with mixed success, with the encouraging preliminary data suggesting pupillometry may provide a metric of threat-related conflict detection being tempered by the lack of trait-related effects observed. This lack of an effect limited a full exploration of the hypotheses informed by the expansive theoretical and methodological considerations described in the introduction. Nevertheless, with several potential avenues identified for improving the method described in the present study, the exploration of the promising theoretical description of narcissistic aggression as increased conflict sensitivity should be considered worthy of further research.

References

Al-Samarraie, H., Sarsam, S. M., Alzahrani, A. I., & Alalwan, N. (2018). Personality and individual differences: The potential of using preferences for visual stimuli to predict the Big Five traits. *Cognition, Technology & Work, 20*(3), 337–349. https://doi.org/10.1007/s10111-018-0470-6

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology, 53*, 27–51. https://doi.org/10.1146/annurev.psych.53.100901.135231

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences, 8*(12), 539–546. https://doi.org/10.1016/j.tics.2004.10.003

Brown, A. A., Freis, S. D., Carroll, P. J., & Arkin, R. M. (2016). Perceived agency mediates the link between the narcissistic subtypes and self-esteem. *Personality and Individual Differences*, *90*, 124–129. https://doi.org/10.1016/j.paid.2015.10.055

Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology*, *63*(3), 452–459. https://doi.org/10.1037/0022-3514.63.3.452

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods*, *40*(1), 109–115. https://doi.org/10.3758/brm.40.1.109

Cascio, C. N., Konrath, S. H., & Falk, E. B. (2015). Narcissists' social pain seen only in the brain. *Social Cognitive and Affective Neuroscience*, *10*(3), 335–341. https://doi.org/10.1093/scan/nsu072

Chester, D. S., & DeWall, C. N. (2016). Sound the Alarm: The Effect of Narcissism on Retaliatory Aggression Is Moderated by dACC Reactivity to Rejection. *Journal of Personality*, *84*(3), 361–368. https://doi.org/10.1111/jopy.12164

De Panfilis, C., Antonucci, C., Meehan, K. B., Cain, N. M., Soliani, A., Marchesi, C., Clarkin, J. F., & Sambataro, F. (2019). Facial Emotion Recognition and Social-Cognitive Correlates of Narcissistic Features. *Journal of Personality Disorders*, *33*(4), 433–449. https://doi.org/10.1521/pedi_2018_32_350

Eisenberger, N. I. (2015). Meta-analytic evidence for the role of the anterior cingulate cortex in social pain. *Social Cognitive and Affective Neuroscience*, *10*(1), 1–2. https://doi.org/10.1093/scan/nsu120

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An FMRI study of social exclusion. *Science (New York, N.Y.)*, *302*(5643), 290–292. https://doi.org/10.1126/science.1089134

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Fenske, M. J., & Eastwood, J. D. (2003). Modulation of focused attention by faces expressing emotion: Evidence from flanker tasks. *Emotion (Washington, D.C.)*, *3*(4), 327–343. https://doi.org/10.1037/1528-3542.3.4.327

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Hart, W., Adams, J. M., & Tortoriello, G. (2017). Narcissistic responses to provocation: An examination of the rage and threatened-egotism accounts. *Personality and Individual Differences*, *106*, 152–156. https://doi.org/10.1016/j.paid.2016.10.049

Heilbronner, S. R., & Hayden, B. Y. (2016). Dorsal Anterior Cingulate Cortex: A Bottom-Up View. *Annual Review of Neuroscience*, *39*, 149–170. https://doi.org/10.1146/annurev-neuro-070815-013952

Hendin, H. M., & Cheek, J. M. (1997). Assessing hypersensitive narcissism: A reexamination of Murray's Narcism Scale. *Journal of Research in Personality*, *31*(4), 588–599. https://doi.org/10.1006/jrpe.1997.2204

Hershman, R., Levin, Y., Tzelgov, J., & Henik, A. (2021). The contribution of meaning to the detection of task conflict. *Quarterly Journal of Experimental Psychology*, *74*(9), 1553–1561. https://doi.org/10.1177/17470218211001331

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. van de. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Horstmann, G., & Bauland, A. (2006). Search asymmetries with real faces: Testing the anger-superiority effect. *Emotion (Washington, D.C.)*, *6*(2), 193–207. https://doi.org/10.1037/1528-3542.6.2.193

Horstmann, G., Borgstedt, K., & Heumann, M. (2006). Flanker effects with faces may depend on perceptual as well as emotional differences. *Emotion (Washington, D.C.)*, *6*(1), 28–39. https://doi.org/10.1037/1528-3542.6.1.28

Houlcroft, L., Bore, M., & Munro, D. (2012). Three faces of Narcissism. *Personality and Individual Differences*, *53*(3), 274–278. https://doi.org/10.1016/j.paid.2012.03.036

Huang, S.-L., Chang, Y.-C., & Chen, Y.-J. (2011). Task-irrelevant angry faces capture attention in visual search while modulated by resources. *Emotion (Washington, D.C.)*, *11*(3), 544–552. https://doi.org/10.1037/a0022763

Hubert Lyall, I., & Järvikivi, J. (2021). Listener's personality traits predict changes in pupil size during auditory language comprehension. *Scientific Reports*, *11*(1), 5443. https://doi.org/10.1038/s41598-021-84886-3

Inzlicht, M., Bartholow, B. D., & Hirsh, J. B. (2015). Emotional foundations of cognitive control. *Trends in Cognitive Sciences*, *19*(3), 126–132. https://doi.org/10.1016/j.tics.2015.01.004

Jauk, E., Benedek, M., Koschutnig, K., Kedia, G., & Neubauer, A. C. (2017). Self-viewing is associated with negative affect rather than reward in highly narcissistic men: An fMRI study. *Scientific Reports*, *7*(1), 5804. https://doi.org/10.1038/s41598-017-03935-y

Kawamoto, T., Onoda, K., Nakashima, K., Nittono, H., Yamaguchi, S., & Ura, M. (2012). Is dorsal anterior cingulate cortex activation in response to social exclusion due to expectancy violation? An fMRI study. *Frontiers in Evolutionary Neuroscience*, *4*, 11. https://doi.org/10.3389/fnevo.2012.00011

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary Stroop effects. *Cognitive Processing*, *12*(1), 13–21. https://doi.org/10.1007/s10339-010-0370-z

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *7*(1), 18–27. https://doi.org/10.1177/1745691611427305

Lawrence, C., & Hutchinson, L. (2013). The influence of individual differences in sensitivity to provocations on provoked aggression. *Aggressive Behavior*, *39*(3), 212–221. https://doi.org/10.1002/ab.21473

Lorber, M. F. (2004). Psychophysiology of aggression, psychopathy, and conduct problems: A meta-analysis. *Psychological Bulletin*, *130*(4), 531–552. https://doi.org/10.1037/0033-2909.130.4.531

Maciantowicz, O., & Zajenkowski, M. (2020). Is narcissistic anger fueled by neuroticism? The relationship between grandiose and vulnerable narcissism,

neuroticism, and trait anger. *Current Psychology*, *39*(5), 1674–1681. https://doi.org/10.1007/s12144-018-9866-y

McDermott, T. J., Wiesman, A. I., Proskovec, A. L., Heinrichs-Graham, E., & Wilson, T. W. (2017). Spatiotemporal Oscillatory Dynamics of Visual Selective Attention during a Flanker Task. *NeuroImage*, *156*, 277–285. https://doi.org/10.1016/j.neuroimage.2017.05.014

Okada, R. (2010). The relationship between vulnerable narcissism and aggression in Japanese undergraduate students. *Personality and Individual Differences*, *49*(2), 113–118. https://doi.org/10.1016/j.paid.2010.03.017

Orr, C., & Hester, R. (2012). Error-related anterior cingulate cortex activity and the prediction of conscious error awareness. *Frontiers in Human Neuroscience*, *6*. https://www.frontiersin.org/article/10.3389/fnhum.2012.00177

Pérez-Dueñas, C., Acosta, A., & Lupiáñez, J. (2014). Reduced habituation to angry faces: Increased attentional capture as to override inhibition of return. *Psychological Research*, *78*(2), 196–208. https://doi.org/10.1007/s00426-013-0493-9

Raine, A. (2002). Annotation: The role of prefrontal deficits, low autonomic arousal, and early health factors in the development of antisocial and aggresive behavior in children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *43*, 417–434. https://doi.org/10.1111/1469-7610.00034

Ridderinkhof, K. R., Wylie, S. A., van den Wildenberg, W. P. M., Bashore, T. R., & van der Molen, M. W. (2021). The arrow of time: Advancing insights into action control from the arrow version of the Eriksen flanker task. *Attention, Perception, & Psychophysics*, *83*(2), 700–721. https://doi.org/10.3758/s13414-020-02167-z

Rogoza, R., Żemojtel-Piotrowska, M., Kwiatkowska, M. M., & Kwiatkowska, K. (2018). The Bright, the Dark, and the Blue Face of Narcissism: The Spectrum of Narcissism in Its Relations to the Metatraits of Personality, Self-Esteem, and the Nomological Network of Shyness, Loneliness, and Empathy. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/article/10.3389/fpsyg.2018.00343

Rotge, J.-Y., Lemogne, C., Hinfray, S., Huguet, P., Grynszpan, O., Tartour, E., George, N., & Fossati, P. (2015). A meta-analysis of the anterior cingulate

contribution to social pain. *Social Cognitive and Affective Neuroscience*, *10*(1), 19–27. https://doi.org/10.1093/scan/nsu110

Sherman, E. D., Miller, J. D., Few, L. R., Campbell, W. K., Widiger, T. A., Crego, C., & Lynam, D. R. (2015). Development of a Short Form of the Five-Factor Narcissism Inventory: The FFNI-SF. *Psychological Assessment*, *27*(3), 1110–1116. https://doi.org/10.1037/pas0000100

Sirois, S., & Brisson, J. (2014). Pupillometry. *WIREs Cognitive Science*, *5*(6), 679–692. https://doi.org/10.1002/wcs.1323

Sleegers, W. W. A., Proulx, T., & van Beest, I. (2017). The social pain of Cyberball: Decreased pupillary reactivity to exclusion cues. *Journal of Experimental Social Psychology*, *69*, 187–200. https://doi.org/10.1016/j.jesp.2016.08.004

Snowden, R. J., O'Farrell, K. R., Burley, D., Erichsen, J. T., Newton, N. V., & Gray, N. S. (2016). The pupil's response to affective pictures: Role of image duration, habituation, and viewing mode. *Psychophysiology*, *53*(8), 1217–1223. https://doi.org/10.1111/psyp.12668

Tannert, S., & Rothermund, K. (2020). Attending to emotional faces in the flanker task: Probably much less automatic than previously assumed. *Emotion (Washington, D.C.)*, *20*(2), 217–235. https://doi.org/10.1037/emo0000538

Unsworth, N., Robison, M. K., & Miller, A. L. (2019). Individual differences in baseline oculometrics: Examining variation in baseline pupil diameter, spontaneous eye blink rate, and fixation stability. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(4), 1074–1093. https://doi.org/10.3758/s13415-019-00709-z

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005–2015. https://doi.org/10.3758/s13423-018-1432-y

Van Steenbergen, H., & Band, G. (2013). Pupil dilation in the Simon task as a marker of conflict processing. *Frontiers in Human Neuroscience*, *7*. https://www.frontiersin.org/article/10.3389/fnhum.2013.00215

Wühr, P., & Heuer, H. (2017). Response Preparation, Response Conflict, and the Effects of Irrelevant Flanker Stimuli. *Advances in Cognitive Psychology*, *13*(1), 70–82. https://doi.org/10.5709/acp-0208-3

Zajenkowski, M., Maciantowicz, O., Szymaniak, K., & Urban, P. (2018). Vulnerable and Grandiose Narcissism Are Differentially Associated With Ability and Trait Emotional Intelligence. *Frontiers in Psychology*, *9*, 1606. https://doi.org/10.3389/fpsyg.2018.01606

Zanarini, M. C., Vujanovic, A. A., Parachini, E. A., Boulanger, J. L., Frankenburg, F. R., & Hennen, J. (2003). A screening measure for BPD: The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD). *Journal of Personality Disorders*, *17*(6), 568–573. https://doi.org/10.1521/pedi.17.6.568.25355

Zarnowski, O., Ziton, S., Holmberg, R., Musto, S., Riegle, S., Van Antwerp, E., & Santos-Nunez, G. (2021). Functional MRI findings in personality disorders: A review. *Journal of Neuroimaging: Official Journal of the American Society of Neuroimaging*, *31*(6), 1049–1066. https://doi.org/10.1111/jon.12924

Appendices

Table A1. Overflow table for non-significant ANOVA results observed in Experiment 1.

	Result
IV: Narcissism, DV: PD	F = 1.13, p = .32
IV: Grandiose narcissism, DV: PD	F = 2.11, p = .12
IV: Vulnerable narcissism, DV: PD	F = .96, p = .38
IV: Narcissism, DV: ACC	F = .45, p = .63
IV: Grandiose narcissism, DV: ACC	F = 2.55, p = .083
IV: Vulnerable narcissism, DV: ACC	F = .16, p = .84
IV: Narcissism, DV: RT	F = 1.46, p = .23
IV: Grandiose narcissism, DV: RT	F = 1.89, p = .15
IV: Vulnerable narcissism, DV: RT	F = .33, p = .71

IV: Narcissism * Congruency, DV: PD	F = .27, p = .59
IV: Grandiose narcissism * Congruency, DV: PD	F = .29, p = .59
IV: Vulnerable narcissism * Congruency, DV: PD	F = .24, p = .62
IV: Narcissism * Congruency, DV: ACC	F = .50, p = .47
IV: Grandiose narcissism * Congruency, DV: ACC	F = .59, p = .44
IV: Vulnerable narcissism * Congruency, DV: ACC	F = .58, p = .44
IV: Narcissism * Congruency, DV: RT	F = .15, p = .69
IV: Grandiose narcissism * Congruency, DV: RT	F = .11, p = .74
IV: Vulnerable narcissism * Congruency, DV: RT	F = .083, p = .77

Table A2. Overflow table for non-significant ANOVA results observed in Experiment 2.

	Result
IV: Narcissism, DV: PD	F = 2.84, p = .064
IV: Grandiose narcissism, DV: PD	F = .54, p = .58
IV: Vulnerable narcissism, DV: PD	F = 2.93, p = .058
IV: Narcissism, DV: ACC	F = .37, p = .68
IV: Grandiose narcissism, DV: ACC	F = .19, p = .82
IV: Vulnerable narcissism, DV: ACC	F = .22, p = .79
IV: Narcissism, DV: RT	F = .57, p = .56
IV: Grandiose narcissism, DV: RT	F = 1.70, p = .18
IV: Vulnerable narcissism, DV: RT	F = 1.85, p = .16
IV: Narcissism * Congruency, DV: PD	F = 1.04, p = .31
IV: Grandiose narcissism * Congruency, DV: PD	F = .88, p = .35
IV: Vulnerable narcissism * Congruency, DV: PD	F = .77, p = .38
IV: Narcissism * Congruency, DV: ACC	F = 1.55, p = .21
IV: Grandiose narcissism * Congruency, DV: ACC	F = .58, p = .55

IV: Vulnerable narcissism * Congruency, DV: ACC	F = 1.53, p = .22
IV: Narcissism * Congruency, DV: RT	F = .37, p = .68
IV: Grandiose narcissism * Congruency, DV: RT	F = 1.11, p = .33
IV: Vulnerable narcissism * Congruency, DV: RT	F = .59, p = .55
IV: Narcissism * Flanker, DV: PD	F = .57, p = .67
IV: Grandiose narcissism * Flanker, DV: PD	F = 2.47, p = .087
IV: Vulnerable narcissism* Flanker, DV: PD	F = .61, p = .65
IV: Narcissism * Flanker, DV: ACC	F = 1.63, p = .16
IV: Grandiose narcissism * Flanker, DV: ACC	F = 1.53, p = .19
IV: Vulnerable narcissism * Flanker, DV: ACC	F = 1.50, p = .20
IV: Narcissism * Flanker, DV: RT	F = 1.76, p = .13
IV: Grandiose narcissism * Flanker, DV: RT	F = 2.39, p = .053
IV: Vulnerable narcissism * Flanker, DV: RT	F = .50, p = .73
IV: Narcissism * Target, DV: PD	F = 1.94, p = .10
IV: Grandiose narcissism * Target, DV: PD	F = 1.46, p = .21
IV: Vulnerable narcissism* Target, DV: PD	F = .16, p = 95
IV: Narcissism * Target, DV: ACC	F = .57. p = .68
IV: Grandiose narcissism * Target, DV: ACC	F = .34, p = .85
IV: Vulnerable narcissism * Target, DV: ACC	F = .45, p = .76
IV: Narcissism * Target, DV: RT	F = .70, p = .59
IV: Grandiose narcissism * Target, DV: RT	F = .64, p = .63
IV: Vulnerable narcissism * Target, DV: RT	F = .70, p = .59

5. Narcissism in-silico: word frequency-derived and qualitative accounts of narcissism and narcissistic aggression in a social online community

5.1. Chapter Overview

The study described in the following chapter represents a significant pivot from the approach taken in the previous chapters, documenting an exploratory study that moves away from lab-based operationalisations of narcissistic aggression in favour of using a naturalistic dataset and a mixed methods approach. As a result, a brief narrative overview is provided. The study utilizes web scraping to extract a large volume of text data consisting of contributions from individual members of a community on the major social media website Reddit, which is analysed using both a qualitative thematic analysis and a quantitative text frequency analysis. This pivot offers an alternative perspective, viewing narcissism in naturalistic and social linguistic behaviours, without the potentially occluding influence of operationalisations of narcissistic aggression relying on self-report or artificial behavioural tasks. This shift in perspective may provide insights into how narcissistic behaviour manifests more organically when influenced by forces that are less present in a laboratory context. Adopting this approach, the study documented in this chapter explores both the potential utility of studying narcissistic aggression as a qualitative phenomena, and the more general utility of studying narcissistic behaviour in naturalistic datasets. The mixed methods aspect of this study builds on the existing literature examining narcissism in naturalistic online datasets with text frequency analysis, attempting to replicate and expand upon these findings by investigating the manifestation of narcissism and narcissistic subtypes in text frequency data. Importantly, this is made possible by taking a synergistic approach where qualitative insights guide the interpretation of text frequency data.

5.2. Introduction

The previous studies in this thesis have explored two broad approaches to studying narcissistic aggression. The first utilised classic psychometric and behavioural psychology methods to operationalise narcissism and aggression, whereas the second experimented with physiological and related behavioural methods intended to be more valid by virtue of not depending on subjective self-report measures. This second approach was met with limited success, suggesting the relatively safer option presented by self-report methods may be favourable despite their limitations. However, this approach emphasising the reduction of narcissistic aggression to an operationalisation in terms of more objective methods is not the only alternative to self-report methods. It is also possible to lean more heavily into operationalising narcissistic aggression as an abstract and non-reducible phenomenon, removing the implicit assumption it is a trait that can be measured objectively, and therefore reducing the relevance of self-report bias.

An alternative approach to studying narcissism

The consequence of taking an alternative approach of this nature is that it necessarily involves a radical change in research methodology, adopting qualitative methods which naturally come with their own practical considerations. As a result, it is not a case of one approach being arbitrarily better than the other, but rather considering which approach is more appropriate given the research objective. A qualitative approach may be able to provide an account of narcissistic aggression in terms of the way individuals more passively (i.e. in the sense they are not compelled to do so by an experimenter) describe their perceptions of themselves and others. That is to say, qualitative approaches provide a view of narcissistic aggression outside of the laboratory, which due to the relative scarcity of research using this approach may provide new insight into how narcissistic behaviour manifests in more naturalistic contexts. Therefore if the objective is to develop a deeper and clearer understanding of narcissistic aggression than is currently available in the research literature, a qualitative approach may offer a useful change of perspective.

Web scraping online communities for naturalistic data

It is important to note that qualitative approaches do not necessarily entail setting aside quantitative methods altogether, as it is possible to apply quantitative methods to the study of naturalistic qualitative datasets, for example using language frequency analyses on dialogue or other types of free text data (e.g. Kern et al., 2014; Yarkoni, 2010). In either case, access to a naturalistic qualitative dataset is necessary. Traditionally, access to datasets of this type has been challenging to achieve, requiring lengthy data collection procedures which put an effective limit on sample size. Whilst the extent to which sample size is relevant in qualitative research is debated (Braun & Clarke, 2021), it nevertheless remains a factor limiting the ability to investigate important heterogeneities that may be present in a dataset. These primarily practical limitations are amplified by the context of a global pandemic within which the present study is conducted. However, the unparalleled modern access to online data and real interactions between individuals provides the potential for a remarkable change to the accessibility of naturalistic datasets, with a diverse selection of active, dynamic social communities consisting of thousands of individuals available to sample. Data from these communities can be collected using web scraping, a term for a variety of techniques whereby large volumes of data are extracted in bulk from legally accessible online databases. Many major websites (e.g. social media platforms such as Twitter) have taken deliberate efforts to facilitate and regulate access to their data via web scraping. It has proved a useful technique in psychological research for collecting large quantities of behavioural data regarding online habits (Landers et al., 2016; Speckmann, 2021), but the utility of web scraping also lies in rapidly collecting rich datasets of text data extracted from written user submissions to websites (such as comments on forum or social media platforms). The length, content and richness of these submissions varies from website to website, but as these user submissions can number in millions on major websites, they represent a massively diverse dataset. Importantly, this data is also naturalistic, consisting largely of thoughts, observations and opinions motivated independently by the individual, as well as organic interactions between individuals or within groups of individuals. As a result, web scraped data of this type provides an ideal and easily accessible dataset for

either of the qualitatively-oriented approaches to studying narcissistic aggression discussed above.

Narcissistic aggression in web scraped naturalistic datasets: potential insights

The overall aim of the present study is presenting an exploratory evaluation of this particular naturalistic approach, but in the process also aims to generate new insight into narcissistic aggression not accessible by the methods documented in previous chapters. Indeed, studying narcissistic aggression using this approach not only offers an alternative methodological perspective, but also allows for new subtype differences in narcissism to be studied that are not as easily assessed outside of these naturalistic contexts. Communal narcissism is a relatively recently defined subtype of narcissism that is particularly relevant to an operationalisation of narcissism in terms of highly social qualitative data (Gebauer et al., 2012; Gebauer & Sedikides, 2018). Communal narcissism, like other subtype traits, is defined by the same core features of the narcissistic personality. In terms of the theoretical framework of narcissism developed in this thesis, these core features are an exaggerated self-perception shaped top-down by narcissistic beliefs, that is constantly in jeopardy of being challenged by bottom-up information concerning observations inconsistent with these narcissistic beliefs (Yang et al., 2018). This underlies a behavioural system that utilises a range of behavioural and cognitive strategies to guard against this threat and reinforce narcissistic beliefs. However, where other narcissistic subtypes may achieve this through highly individualistic strategies (e.g. directly signalling positive qualities or minimising negative qualities of the individual), communal narcissism emphasises (perceived) prosocial strategies that guard and reinforce narcissistic self-perceptions by making them self-evident in an individual's acceptance and approval within a community (Gebauer et al., 2012). As a result, while communal narcissism is similarly associated with self-interested attempts to influence conditions in a social group to be more amenable to narcissistic beliefs (e.g. increasing the acceptance and endorsement of superficial narcissistic signalling strategies), these attempts are intended to be perceived overtly as agreeable and in accordance with the rules of a community (Joyce et al., 2019; Rentzsch & Gebauer, 2019).

Where grandiose and vulnerable narcissism represent one dimension of subtype differences, communal narcissism is located on another subtype dimension opposed (at least superficially) to agentic narcissism, which represents a more classical individualistic subtype less concerned with community approval (Gebauer et al., 2012). Communal and agentic subtype differences also highlight the ways in which narcissistic traits may not be constant but rather adapt to accommodate situational factors (Grapsas et al., 2020). That is to say, whilst some individuals may be more disposed to communal or agentic forms of narcissistic behaviour than others, it may also be the case that individual differences in these traits are determined by the extent to which either communal or agentic behaviour is facilitated by the context. For example, for a given narcissistic individual, communal narcissistic traits may be emphasised in social contexts where individual control is difficult but can be gained indirectly through pursuing community approval, whilst agentic traits may be emphasised in social contexts where prosocial behaviour is less necessary to assert influence (Denissen et al., 2013; Grapsas et al., 2020).

Selecting an appropriate online community for studying narcissistic aggression

Major online communities are highly social and typically feature a large degree of bottom-up control of behaviour through relying on community acceptance to achieve popularity and avoid censorship, and therefore provide a highly appropriate source of data for investigating communal narcissism, and the influence of social situational factors on narcissistic behaviour more generally (see Balcerowska & Sawicki, 2022; Buffardi & Campbell, 2008; Joyce et al., 2019; Kristinsdottir et al., 2021). Web scraped data provides an effective means of examining the behaviour of individuals within these communities, granting access to a large population of individuals that can be sampled in order to maximize the variance of individual differences in narcissistic traits, whilst also allowing the degree of specificity necessary to analyse the particular narcissistic behaviours of each individual. Moreover, as online communities (specifically social media communities) increasingly come to define and dictate many modern forms of communal social activity, studying narcissism in these

contexts may provide a more representative insight into how narcissistic behaviour manifests in contemporary social communities.

In the present study, a mixed methods approach is taken to investigate narcissistic aggression in user submissions collected from a large online community, both to explore the ways in which narcissistic aggression manifests in the linguistic devices of individuals within these communities, and validate the usefulness of this approach (i.e. in terms of whether this approach yields insights beyond the scope of what is already known). Specifically, this mixed methods approach will entail both a qualitatively-oriented method in the form of a deductive semantic thematic analysis (Braun et al., 2019) and a quantitatively-oriented method in the form of text frequency analysis (Kennedy et al., 2021). These provide two distinct but compatible means of investigating the same naturalistic qualitative dataset. In the present study, this dataset will be collected through web scraping the major social media website Reddit. Whilst other social media websites that emphasise more superficial forms of user contribution and social interaction via photo media may be more naturally accustomed to some forms of narcissism, Reddit is considered to be a more appropriate data source for the present study for several reasons. Contributions by Reddit users cannot be traced to an actual individual. Instead, each contributor has a username, and anyone on the website can view content contributed to Reddit. This makes Reddit a useful data-source for the purposes of the proposed study, as users provide rich data, not limited by length, without the social desirability bias that might be expected on websites with some connection to their real identity. Furthermore, Reddit is well suited to investigating interpersonal behaviour, including aggressive behaviour: beyond being simply a platform for users to express their views, it is a platform where users can express their views in response to the contributions of others, both for purposes of support, debate or attack.

The former qualitatively-oriented approach is relatively novel, insofar as there has been little to no investigation of narcissism using this style of thematic analysis, and so the theoretical account of narcissism and narcissistic aggression provided earlier in this thesis (see **Chapter 1**) will provide a framework. The latter quantitatively-

oriented approach using text analysis has been used in some previous research (Ireland & Mehl, 2014; Kern et al., 2014; Yarkoni, 2010), for example a word frequency analysis investigating language use associated with grandiose narcissism found grandiose narcissism was associated with an increased frequency of second-person pronouns, sex and sexuality related words and swear words; and a decreased frequency of tentative words, anxiety or fear related words, and words related to senses and sensation (Holtzman et al., 2019). As a result it is possible to use the linguistic metric of grandiose narcissism identified by this research as a means of measuring grandiose narcissism in the web scraped dataset featured in the present study, where other quantitative measures of narcissistic traits in the sample are not available. Language use associated with aggression can be measured more straightforwardly using word frequency metrics that are typically default integrated in text analysis software packages (Pennebaker et al., 2015). However, there is little to no pre-existing psychological research of either kind using Reddit data.

Approaches and aims of the present study

In summary, investigating narcissistic aggression using a naturalistic and qualitative dataset may provide a useful alternative approach that avoids the shortcomings of more rigid laboratory-based and quantitative approaches, whilst also contributing to an understanding of how narcissism is evidenced organically in the autonomous linguistic behaviour of individuals within the online communities that dominate the contemporary social environment. A qualitatively-oriented thematic analysis method has the potential to yield insightful analysis into the content of this behaviour (i.e. the ways in which theoretical features of narcissism and narcissistic subtypes manifest in the data). Alongside this, a quantitatively-oriented text analysis method retains the search for viable quantitative operationalisations of narcissistic aggression that has driven the studies documented in this thesis. Importantly, this is done whilst not engaging in the strongly reductive approach of the previous study by grounding this method in a naturalistic qualitative dataset.

The present study adopts these methods as two distinct but related approaches. The first is an open-ended exploration of linguistic devices relating to narcissistic

aggression in user contributions to Reddit. Driven by an existing theoretical framework of narcissism (as a defensive set of strategies intended to protect and maintain a fragile sense of self-esteem; see **Chapter 1**), the aim of this exploration is to examine how theoretical concepts in the framework manifest in the language featured in user contributions, which features of the framework (if any) are emphasised in communal contexts, and ultimately update the framework based on these observations. It is anticipated that communal motivations to conform and contribute appropriately will result in narcissistic aggression becoming narrowed into a small number of socially acceptable behaviours (Gebauer & Sedikides, 2018; Joyce et al., 2019; Rentzsch & Gebauer, 2019), which are less superficially obvious and agentic than classical narcissistic behaviours but nevertheless able to be evidenced and characterised by subtype differences (Brailovskaia et al., 2020). For example, it is anticipated that grandiose narcissism will manifest as an emphasis on passively signalling superiority (Grapsas et al., 2020) and beliefs and behaviours that are more extraverted or strongly expressed (Jauk et al., 2017). In contrast, it is anticipated that vulnerable narcissism will manifest as an emphasis on cynical and avoidant beliefs and behaviours, reflecting increased hostility and neuroticism, and signalling disgust or disapproval of lower status group members or others when permissible (Miller et al., 2011; Zeigler-Hill et al., 2021).

The second is an investigation of narcissistic aggression in terms of word frequency. Word frequency data generated via text analysis of user contributions to Reddit relating to aggression (using a default word frequency metric aggressive language) and grandiose narcissism (using word frequency metrics identified by prior research) will be compared with assessments of narcissistic traits identified via thematic analysis. For the purposes of the more quantitatively-oriented approach used in this aspect of the study, the results of the thematic analysis will be represented here as a synthesis of both a qualitative assessment of the extent to which each participant (i.e. a user included in the final web scraped sample) exhibits narcissistic traits based on the results of the thematic analysis, and a quantitative metric of this assessment in the form of 'code frequency', which refers to the quantity of codes assigned to each

participant as part of the thematic analysis (see **Discussion** for a defense of the use of code frequency as a metric).

More generally, the aim of this synthesis of quantitative and qualitative data (for examples of related approaches see; Grayman-Simpson, 2009; Paramboukis et al., 2016; Utama et al., 2020) is to allow the results of the text analysis to be both contrasted and correlated with the results of the qualitative analysis. The intention of this contrast is not to render the qualitative thematic analysis redundant through arbitrary reduction to a quantitative measure, but rather to allow the thematic analysis results to drive the development of new metrics from text analysis data, and validate previous research findings. This synthesis allows text analysis data and thematic analysis data to be correlated to discover new relationships between word frequency and narcissistic subtypes, and validate whether previously identified word frequency metrics of grandiose narcissism established by previous research are coherent with qualitative assessments of grandiose narcissism. It is anticipated that the narcissistic traits identified in the thematic analysis will be associated with corresponding word frequency metrics, which based on previous research may be informal language and use of swear words, and/or language relating to sex and sexuality, negative emotion, and/or anger (Bogolyubova et al., 2018; Ireland & Mehl, 2014). However, given the broader scope of the theoretical framework used in the analysis, word frequency are also anticipated to reflect more specific linguistic behaviours associated with the hypothesised differences in subtype traits (e.g. anxiety, tentativeness, reward/risk focus, etc.).

5.3. Method

Participants

Reddit is divided into a large number of sub-communities of varying size, each catered to the discussion of a specific topic. These communities are referred to as ‘subreddits’. The variety of subreddits across Reddit is extremely diverse, and naturally the topic focus of a subreddit has a considerable influence on the type of

content submitted by users. As a result, it was necessary to select a specific subreddit or a limited number of subreddits to web scrape, with the choice being appropriate to the research question (see below). Furthermore, each subreddit is controlled and moderated by a small number of users from within the community, who to some extent are free to enforce their own rules governing what content and behaviour is acceptable. As a result, when collecting a sample of users via web scraping, it was necessary to take a deliberate and selective approach that takes into account not only the topic, but also the particular culture of the target subreddit. Finally, as the most popular subreddits (and therefore likely the best targets for web scraping given the potential richness of data) can have a very large population of diverse users, producing a massive quantity of content, it was necessary to identify a sample that is constrained given the practical limits of available time and resources, whilst maintaining as much richness of user data as possible

The present study examined contributions specifically within the subreddit 'r/relationships'. Given the highly exploratory nature of the study, there was little to no literature to draw on to inform this choice, and there are numerous alternative communities that may be similarly relevant to the study of narcissism and narcissistic aggression. Nevertheless, this subreddit was deemed to be most appropriate for the purposes of the present study based on the following rationale; (1) it is popular ensuring a large population to sample from, (2) it is based on giving advice and debating issues and therefore strongly emphasises interpersonal content (i.e. directly inviting interaction from other users), ensuring data is rich in social interactions between users, (3) it is relatively uncontrolled in terms of submission format and therefore conducive to users behaving freely, reducing the influence of overbearing community norms that may limit or bias data, (4) due to its emphasis on advice and discussion it is typically host to disagreement and exchange of opinions between users, making it an appropriate dataset for investigating aggressive online behaviours, (5) due to its emphasis on the evaluation and discussion of the relationships and lives of others, it features many opportunities for users to engage in both explicit and implicit narcissistic social signalling, the criticism or shaming of others, and the expression of narcissistic beliefs, making it an appropriate dataset for investigating

narcissism, and (6) it likely attracts a diverse population of users given its widely applicable topic, ensuring individual differences in personality traits are likely to be sufficiently variable. Importantly, the 'r/relationships' subreddit encourages both content that is explicitly prosocial (i.e. caring and supportive), and content that is implicitly prosocial but overtly features criticism that is intended to be in some way constructive (although naturally the interpretation of what constitutes constructive criticism is subjective and certainly influenced by trait differences, including narcissism). This means that user submissions are not critical by necessity, and whether a user engages in either supportive behaviour or critical behaviour is optional. This in turn means that the more critical and hostile behaviours hypothesised to be associated with narcissism represent a meaningful behavioural choice.

The experiment was approved by the School of Psychology ethics committee (Ref: S1267).

Data collection

Data-collection was performed using a custom web scraping algorithm. Web scraping is more precisely defined as a method used to collect data stored on a website that is accessible through a website's API (application programming interface), a gatekeeping system that allows external users to request data from a website in accordance with rules specified by the API. Web scraping relies on code that makes many thousands of specific requests from the API automatically, resulting in a large volume of information being rapidly pulled and exported to a local database. Reddit allows batch downloading of data via its API via web scraping. Accessing and analysing Reddit data using this method is fully legal and permitted by the website's terms and conditions.

The web scraping script used in the present study was programmed in Python3, and was able to collect all submissions within a specified subreddit within a given time period, including data attached to each submission or comment, organised by the user that contributed the submission. The time period used was a four month time period

between the 1st of January and the 1st of May 2020. This limited duration was selected as the very high volume of content submitted to the subreddit in even a single month renders analysing any dataset produced with more than four months worth of data impractical (i.e. four months generated ~400MB of raw text data).

Data preprocessing was performed to refine the sample to consist only of data associated with the 50 most active users. This was both to produce a final sample consisting only of users with rich sets of associated data, and to exclude the vast majority of users who were not viable for inclusion in the sample due to only making a very small number of contributions to the subreddit, whilst maintaining a large enough sample size to allow for individual differences.

Data Analysis

Text analysis

The dataset was first analysed using LIWC2015 (Pennebaker et al., 2015), a software package which measures word frequencies within a text dataset, and then assigns individual word frequencies to broader word frequency categories constituting several hundred words and organised by a shared topic and high internal consistency (i.e. correlations between word use). For example, the individual frequencies of the words 'hate', 'kill' and 'annoyed' are aggregated (along with many others) into the broader 'anger' word frequency category. This translates the bulk qualitative dataset associated with each participant into a quantitative dataset of aggregated word frequency statistics.

Given the lack of prior research to inform expectations regarding what might constitute interesting and/or relevant individual differences in word frequency in our sample, this text analysis was largely performed for the purpose of later synthesis with the results of the qualitative analysis described below. However, the exception was the word frequency metric of grandiose narcissism identified by previous research (Holtzman et al., 2019) and described above. As this metric was also identified using the LIWC software package, it could be applied directly to the

transformed dataset in the present study to establish individual differences. Specifically, this metric consists of increased frequency of the categories 'second-person pronouns', 'sex and sexuality' related words and 'swear' words, and a decreased frequency of the categories 'tentative' words, 'anxiety or fear' related words, and words related to 'senses and sensation'. By appropriately combining the frequency of these categories, a merged variable can be generated representing language use indicative of grandiose narcissism (i.e. individuals with higher or lower frequencies in the respective categories will be associated with higher grandiose narcissism).

Using this metric, individuals in the refined sample of the 50 most active users were sorted from high to low based on their word frequency-derived trait grandiose narcissism score. A final refined sample was then generated from the five highest and five lowest scoring users, in order to generate a sample for thematic analysis with the maximum possible variation in trait narcissism. This was necessary to ensure that the sample actually features meaningful individual differences in narcissism, to facilitate the identification and exploration of narcissistic behaviour. It is important to acknowledge that grandiose narcissism and narcissism are not equivalent, and so it was used as a means of selecting the final sample only due to the absence of any other means of measuring a more general trait narcissism. It is assumed that although these individuals were more likely to be characterised by grandiose narcissism (or a lack thereof), individual differences in this subtype should be to some extent associated with individual differences in related subtypes. For example, narcissistic grandiosity should not necessarily preclude behaviours that are characterised by vulnerability. Regardless, it should be stressed that this exclusively grandiose metric of narcissism is only featured here out of necessity, and in contrast the subsequent thematic analysis features an examination and consideration of all narcissistic subtypes. Finally, to blind the analysis moving forward and prevent researcher knowledge of which users in the final sample were rated as high or low in grandiose narcissism, the selected users were anonymised via the replacement of their username with a random ID string.

Thematic analysis

The final refined dataset were then analysed using thematic analysis. Thematic analysis is a qualitative analysis technique wherein features of the text being analysed are categorised as codes, which are then grouped into themes that represent broader concepts. The process is recursive, and codes and themes are regularly revised in order to improve consistency and limit redundancy. A deductive, semantic-level approach was taken to coding the data and identifying themes in the present study (Braun et al., 2019). In practical terms, this means the analysis was driven by a pre-existing theoretical framework that guided how the language used by individuals in the dataset was interpreted and classified (deductive), but focused primarily on the superficial meaning of language used without substituting an alternative deeper meaning (semantic). For example, the phrase 'I can't believe you have such low standards!' is identified as potentially narcissistic based on deductive reasoning, but is classified as (e.g.) 'shaming others choices' based on the semantic features of the phrase, and not as (e.g.) 'deflecting from personal insecurities' which relies on interpretation of latent features. The specific framework driving the deductive approach is the theoretical description of narcissism and narcissistic aggression outlined in **Chapter 1**. Broadly, based on this framework, narcissism is anticipated to manifest as behaviours indicating any of the following: signalling of positive attributes, superior attitudes towards others based on perceived positive attributes, the humiliation or criticism of others, strong language and verbal aggression particularly in reaction to perceived or experienced provocation, hostile attributions and cynical views regarding the intent of others and expected outcomes, feelings of anxiety and insecurity in the context of interpersonal situations, recommendation of excessive confrontational or avoidant behaviour, unempathetic or callous behaviour, described instability in lifestyle and interpersonal relations, insecure attachments, and negative affective language. Importantly, the deeper epistemological assumption implicit to this theoretical framework is critical realism, wherein it is assumed that some objective and directly causal trait phenomenon of narcissism exists, but the observed manifestation of narcissism is necessarily mediated by the methods used to study it. The objective of the thematic analysis was to identify language use indicative of

narcissism, including both general trait features of narcissism and vulnerable, grandiose, communal and agentic subtype trait features. General trait features were taken to be linguistic behaviours exhibiting narcissistic social signalling strategies not clearly characterized by a distinct subtype. Linguistic behaviour indicative of neuroticism, hostility, and/or avoidance were taken to represent vulnerable narcissism, whereas grandiose narcissism was taken to be represented by linguistic behaviour that was markedly extraverted and assertive. Communal and agentic features were less straightforward, but were taken to be linguistic behaviours indicative of narcissistic strategies used for superficially prosocial or superficially self-interested purposes respectively.

In addition to understanding how narcissism and narcissistic aggression manifest as behaviour in social online communities, a goal of the analysis was also to develop an account of each user featured in the sample that describes the extent to which they can be characterised as a narcissistic persona, and any narcissistic subtypes emphasised (including specifically any suggested disposition towards narcissistic aggression). This is both a qualitative account (i.e. an assessment based on the themes characterising the dataset of each user) and a quantitative account in terms of code frequency (e.g. the raw number of codes assigned to a user, with more codes within a given theme representing higher magnitude of whatever the narcissistic trait theme represents). Coding was performed by the researcher, with all codes being evaluated by a second and independent trained analyst to ensure the analysis was robust and unbiased. The thematic analysis procedure was facilitated by the use of NVivo 12 Pro.

Synthesis of quantitative and qualitative analyses

The final stage of the analysis was two different syntheses of the qualitative thematic analysis and quantitative text analysis datasets. The first was a qualitatively-oriented synthesis, where each user in the final sample was categorised by their word frequency-derived grandiose narcissism score, and then evaluated in terms of how accurately this sorting reflected their individual qualitative assessments of narcissistic traits (i.e. to what extent did users with high word frequency-derived grandiose

narcissism exhibit a qualitatively described narcissistic persona), and in particular any grandiose narcissistic traits identified by the thematic analysis. The purpose of this first qualitatively-oriented synthesis was to examine whether word frequency-derived and qualitatively-derived descriptions of grandiose narcissism were coherent, with the aim of both validating the previously identified word frequency metric of grandiose narcissism, and highlighting the distinctions (if any) between these two descriptions.

The second was a quantitatively-oriented synthesis, where codes identified in the thematic analysis were converted into a quantitative measure in terms of code frequency (as described above; see **Discussion** for justification of this approach) and categorised via their respective theme (i.e. several code frequencies were aggregated into a single theme frequency statistic). Code frequencies were then entered into a zero order correlation matrix with word frequency categories generated by the text analysis. The purpose of this synthesis was to: (1) identify novel word frequency metrics of narcissistic traits, expanding the scope of available metrics beyond grandiose narcissism, and (2) further validate the previously identified word frequency metric of grandiose narcissism by investigating how this metric correlates with relevant code frequency metrics.

5.4. Results

5.4.i. Text analysis

All 90 of LIWC word frequencies generated by the analysis were non-zero, indicating all users made use of language within each of LIWC's dictionary categories at least once. Excluding the categories relating to basic grammatical features of the text (which may be relevant in the context of exploring relationships between language use and narcissism, but are considered irrelevant in a general descriptive sense), the top five most frequently used word frequency categories were 'social processes' (M = 20.01), 'present focus' (M = 16.44), 'cognitive processes' (M = 15.87), 'relativity' (i.e. space and time; M = 11.47) and 'drives' (i.e. motivational drives; M = 8.39), whereas the five least frequently used word frequency categories were 'death' (M =

.06), ‘religion’ (M = .10), ‘ingestion’ (M = .19), ‘swear words’ (M = .28) and ‘sex and sexuality’ (M = .29).

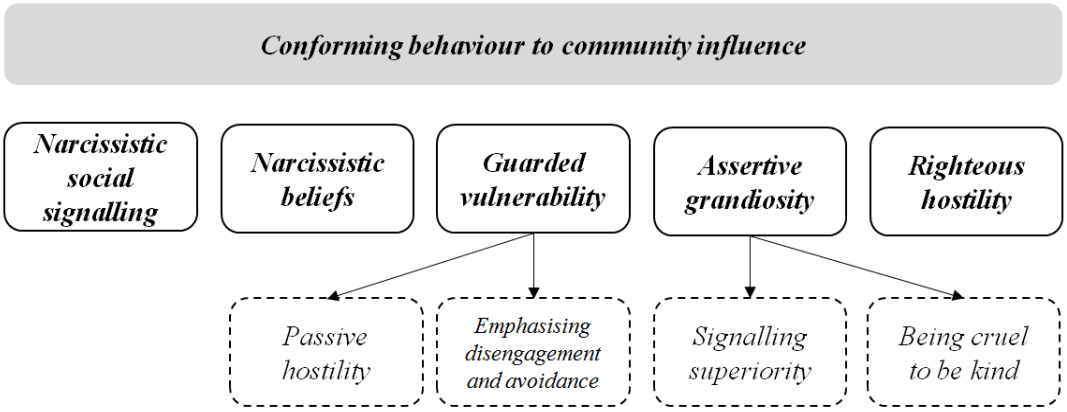
Based on the results of Holtzman et al. (2019) described above, a word frequency metric of grandiose narcissism was generated through dividing the sum of the 'second-person pronouns', 'sex and sexuality' and 'swear words' categories by the sum of the ‘tentative’, ‘anxiety or fear’ and ‘senses and sensation’ categories (i.e. users with the highest sum of the former and lowest sum of the latter having the highest scores on this metric). The mean word frequency-derived grandiose narcissism score was 2.53 (SD = .27) in the high scoring group (consisting of the five highest scorers) and 1.20 (SD = .24) in the low scoring group (consisting of the five lowest scorers). Cronbach’s alpha for the word frequency-derived grandiose narcissism metric was 0.240.

5.4.ii. Thematic analysis

The thematic analysis of narcissistic language use identified five themes relating to the manifestation of narcissistic traits in online social behaviours (see **Figure 1**). These themes were: ‘*Narcissistic social signalling*’, ‘*Narcissistic beliefs’*, ‘*Guarded vulnerability*’, ‘*Assertive grandiosity*’, and ‘*Righteous hostility*’. ‘*Narcissistic social signalling*’ and ‘*Narcissistic beliefs*’ represent the manifestation of non-subtype specific behaviours and beliefs. ‘*Guarded vulnerability*’ represents manifestations of vulnerable narcissism and consists of sub-themes that emphasise passive hostility and avoidance. ‘*Assertive grandiosity*’ represents manifestations of grandiose narcissism and consists of sub-themes relating to controlling behaviour and superiority. Finally, ‘*Righteous hostility*’ represents typically reactive forms of verbal aggression implicitly justified by the context of the social interaction. Although not identified as a discrete theme or subtheme, communal narcissistic traits manifested as a motivational force moderating the manifestation of narcissistic traits across themes. That is to say, communal narcissism was most evident as a behavioural framework for effectively using narcissistic behavioural strategies given the context of community rules and culture. Communal narcissism could therefore be described by

the analysis as a 'meta-theme' (a feature that may be implicit in the raw text data, but emerges more explicitly at the level of themes), titled '*Conforming behaviour to community influence*', the more specific description of which is dependent on the particular form of community influence evident in each theme (described below).

Figure 1. Thematic map showing themes (solid outline) and sub-themes (dashed outline) generated by the thematic analysis. Relationships between a theme and its subthemes are indicated by directional arrows. The 'meta-theme' relating to communal narcissism is indicated in grey.



Narcissistic social signalling

This theme encompasses covert attempts to signal perceived positive qualities associated with narcissistic self-beliefs, through emphasising the shortcomings and failures of others within the community. This may theoretically serve to enhance the effective social status of the individual through lowering the social status of others, or otherwise drawing attention to the apparent gap in status between themselves and another. Linguistic behaviours encompassed by this theme included reprimanding others and expressing explicit disapproval of their actions or beliefs.

> *"Really? After saying it was the most abusive relationship you'd ever seen you recommend this woman stays in hers?"*

Most frequently, this theme included attempts to shame others, expressing disgust and disbelief at the actions or beliefs of others. This behaviour tacitly indicated

disapproval, but more explicitly emphasised the suggestion that an individual should not only act or think differently but feel ashamed about not having acted or thought differently previously.

> *"Wow this guy has no redeeming features and you're sad you broke up? You need higher levels of self esteem."*

Narcissistic beliefs

This theme comprised beliefs expressed explicitly that conformed with narcissistic beliefs predicted by the theoretical framework. These beliefs were consistent with foundational narcissistic beliefs that were not oriented more specifically with a subtype narcissistic trait. In order to ensure conformity with community norms, these beliefs were typically situated in the context of another individual (e.g. beliefs regarding what another individual should do). Expressing these beliefs may have been a deliberate action in order to shift community norms towards becoming more facilitative of narcissistic beliefs, reducing the threat of a narcissistic individual confronting beliefs that are incoherent with their internal narcissistic perception. Alternatively, expressing these beliefs may have simply represented a manifestation of narcissistic beliefs that are usually more closely guarded to appear agreeable, but which are more readily publicly shared when supported by community attitudes, that encourage and protect sharing of beliefs in the context of sharing relevant advice. Beliefs encompassed by this theme included notions of self-importance (i.e. literally the importance of oneself over others) and emphasising prioritising personal needs over the needs of others. Importantly, the others in this context were frequently friends, family members or romantic partners, suggesting self-centred motivation that is distinct from the relatively normal attitude to value one's own needs over those of strangers.

> *"You never owe anyone an explanation of why you are ending a relationship. 'I'm no longer into it,' is the only explanation you need."*

A frequently encountered belief was also the importance of ignoring negativity and not allowing oneself to be influenced or occupied by the distress and problems of others, suggesting a lack of empathy distinct from self-centred beliefs.

> *"Stop worrying about what's good for him. What's good for him is his responsibility, not yours."*

> *"Do what you want. To hell with people's opinions."*

Finally, beliefs in this theme also featured attitudes condoning the use of manipulation and deceit in order to manage relationships with others, reflecting a similarly self-centred and unempathetic attitude but applied more specifically to a behavioural strategy. This recommendation was often justified by being for the express benefit of the recipient, or by being in the interests of the greater good (i.e. manipulating others ultimately for the wellbeing of others).

> *"Do not tell her. Do not tell her. Do not tell her. I know that there is a philosophy out there that says we 'must' be honest about our feelings blah blah blah."*

Guarded vulnerability

This theme constitutes linguistic behaviours and beliefs that indicated a sceptical and aversive attitude towards relationships with others that was frequently defensive in tone. These are coherent with a theoretical description of narcissistic vulnerability, as they both reflect and serve to maintain hostile attitudes towards others, and a predisposition towards avoidance rather than confrontation. This theme is divided into two sub-themes:

Passive hostility. This sub-theme constitutes linguistic behaviours signalling a passive hostility and distrust of others, and expressions of perceived threat and negative intent more generally. These behaviours frequently manifested as typical passive aggression, with users making statements that contained clear critical or negative suggestions that could be legitimately interpreted as aggressive by the

recipient, but that did not contain any explicit indication that they were intended as aggressive by the user.

> *"You can feel how ever you want. Is it possible that you might have ADD?"*

Hostile attitudes were expressed as cynical and wary views, warning other users of imminent or future danger based on perceived cues. These cues were frequently referred to colloquially by community members as 'red flags', indicating stereotyped beliefs regarding predictable covert negative qualities and intentions of others. Cues did not always relate to individuals specifically but could also be situational, indicating future failure or negative outcomes independent of individual actions, and suggestive of a form of cynicism characterised by pessimism and neuroticism.

> *"He doesn't sound over his ex and it sounds like he'd leave you for her in a second if the opportunity ever arose. Do with that info what you will."*

> *"You're ignoring the red flags. Good luck! Update us in a couple of years. He doesn't want to be with you & wants to to [sic] try someone else & see if it works out."*

Emphasising disengagement and avoidance. This sub-theme encompasses attitudes condoning and encouraging disengaging from challenging situations and avoiding conflict where possible. Advice expressing these beliefs typically featured ignoring or blocking contact with individuals associated with the recipient, avoiding investing time or effort in relationships featuring challenge, and ending relationships with issues rather than addressing or confronting those issues. Importantly, as any linguistic behaviours included in this sub-theme reflect a user choosing to engage (i.e. contribute content to the community) rather than disengage, they do not reflect active avoidance on the part of the user themselves, but rather reflect the user expressing beliefs that are suggestive of a disposition towards avoidance.

> *"Always ignore crazy. Don't respond or apologize. Just block them both."*

"I suggest leaving her to her life and finding someone else. Only she can fix herself."

Assertive grandiosity

This theme encompasses linguistic behaviours and beliefs that indicated a superior and callous attitude towards other community members that was typically framed as a well-intentioned desire to help. These are coherent with a theoretical description of narcissistic grandiosity, as they actively assert grandiose narcissistic self-beliefs and position other community members as lower status, and suggest a predisposition towards confrontation. This theme is divided into two sub-themes:

Being cruel to be kind. Adhering with community norms meant otherwise callous and assertive grandiose behaviours frequently manifested in the context of providing advice, where they were justified superficially as attempts to confront a problem head-on and impart wisdom that was blunt and unambiguous. In other words, grandiose behaviours that may have been considered intimidating or antisocial in other contexts were presented as prosocial in a communal context through a 'cruel to be kind' rationale.

"You think being in touch with him will transform him into a man who is into you? This guy is not in love with you, period. You should have walked away the first time."

Advice was frequently callous and unemotional in tone, apparently intended to draw attention to what the user identified as a hard truth, without any attempt to 'soften the blow' through compassionate language and without consideration of the emotional impact advice might have upon the recipient.

"You need a new gf [girlfriend] this one is broken."

"Your hope is naive and fruitless and you need to accept this isn't the woman that you will build a life with."

Furthermore, advice was also strongly directive in nature, often framed less as a suggestion and more of an instruction, without the implication that there may be an alternative interpretation or course of action. This advice also frequently emphasised confrontation through delivering an ultimatum or otherwise assertively tackling a problem.

> *"Get out and break up, in that order. Do NOT talk to him first!!!"*

> *"Open your own bank account ASAP. At a different bank, just in case."*

Signalling superiority. Overtly signalling superiority through the direct expression of positive personal qualities or achievements was a regular feature of this theme. Given the need to conform with community norms, this signalling naturally did not occur independently of an appropriate context, but rather was framed as superficially helpful contextual information to support a point made in advice or better illustrate the problem at hand.

> *"I moved out at 18, 5 years younger than you. Become the adult you're supposed to be and start living your life for yourself."*

Righteous hostility

This theme constitutes linguistic behaviours that explicitly signalled aggressive and critical attitudes towards another individual. These behaviours were constrained and limited by community norms explicitly preventing inappropriate and overtly aggressive behaviour, and as a result were often relatively mild in content. However, the behaviours were also justified by community norms that allow criticism where invited by a recipient, and implicitly allow harsher forms of criticism when a recipient behaves badly or admits to reproachable behaviour or beliefs. As a result, these behaviours were frequently righteous or reprimanding in tone or content, with recipients willingly (but perhaps not wittingly) having invited or justified attacks against them on the basis of their behaviour or beliefs. Notably, behaviours encompassed by this theme were more explicitly focused on direct criticism (i.e. an explicit negative assessment in terms of an individual's behaviour or beliefs) rather

than humiliation (i.e. a broader negative assessment of an individual based on their behaviour or beliefs). Criticism was either 'first degree' and directed at the recipient themselves, or 'second degree' and directed at another individual, which was typically someone mentioned by the recipient or otherwise associated with the recipient.

> *"Your actions were very bad, you were emotionally cheating on your husband. It's no wonder he doesn't trust you anymore."*

> *"The only one who is incompetent here is your bf [boyfriend], who is clearly very socially incompetent."*

This theme also encompasses linguistic behaviours that were more severe than criticism and more closely resembled verbal aggression. In this context, verbal aggression was taken to constitute language intended to be shocking and/or harmful as well as simply critical. Users occasionally used language that was strong or dramatic in character, possibly in excess of what was necessary or appropriate, or used language that was directly insulting. However, the severity of this was variable, and was naturally influenced not only by the individual but also the content and context of the interaction between a user and recipient.

> *"F*ck that noise OP [original poster]. If you want to get back together, then it's on you to get over yourself."*

5.4.iii. Synthesis of thematic analysis and text analysis

Coherence of qualitative and word frequency-derived accounts of narcissism

The results of the thematic analysis were generated using contributions from all users included in the final sample (i.e. the final set of themes and codes describe the dataset independently of any individual in the dataset) and therefore inherently reflect manifestations of narcissistic behaviour shared across the sample. However, despite this, examining the results of the thematic analysis between users revealed clear individual differences in the extent to which particular themes were emphasised in the

behaviour of each user. These differences were most evident between users in the high and low grandiose narcissism groups of the sample (as defined by the word frequency-derived metric of grandiose narcissism).

Table 1. Qualitative accounts of each user and corresponding word frequency-derived grandiose narcissism score for both the high scoring (Users A – E) and low scoring (Users F – J) grandiose narcissism groups.

User ID	Qualitative account of narcissistic traits	GN
A	Shaming, emphasising disengagement, cynical beliefs	2.87
B	Shaming, emphasising disengagement, cynical beliefs	2.75
C	Emphasising disengagement, shaming, strong language	2.43
D	No clearly evident narcissistic traits	2.34
E	Shaming, hostility strong language, cynical beliefs	2.25
F	No clearly evident narcissistic traits	1.34
G	No clearly evident narcissistic traits	1.34
H	Emphasising disengagement, self-centred beliefs	1.30
I	No clearly evident narcissistic traits	1.26
J	Shaming, emphasising disengagement, hostility	.78

** GN: word frequency-derived grandiose narcissism score*

The approach taken for characterising each user in terms of their individually assigned set of codes and themes generated by the thematic analysis was similar to the thematic analysis itself. That is to say, the approach was semantic in that it was based on a straightforward account of which themes were most frequently associated with each user, without an attempt to re-interpret this combination of themes as something other than the sum of its parts. Likewise, the approach was deductive in that the interpretation of this straightforward account was informed by the theoretical framework used in the thematic analysis. It was also naturally informed by the results

of the thematic analysis itself, as the themes and sub-themes encompassing codes naturally dictate how these codes should be interpreted (e.g. a code belonging to the 'guarded vulnerability' theme should be interpreted to represent narcissistic vulnerability).

Qualitative accounts of each user and their respective word frequency-derived grandiose narcissism scores are presented in **Table 1**. Users in the high grandiose narcissism group were characterised primarily by similarities in emphasised themes, along with some notable differences. The most pronounced similarity between users was a strong emphasis on narcissistic signalling behaviour, specifically shaming others, that characterised the high grandiose narcissism group. Users **A**, **B** and **C** were also characterised by a similar strong disposition towards emphasising disengagement and avoidance, which was less evident for users **D** and **E**. Instead, these users were characterised by a shared disposition towards hostility that was distinct from other users. Further differences within the high narcissism group were evident between users **A**, **B** and **E** and the rest of the group. These users were characterised by hostile attitudes that were more overtly vulnerable in nature, with an increased disposition towards pessimistic and distrustful hostile attitudes. User **D** was distinct from the rest of the group due to a markedly lower frequency of codes (i.e. data saturation was reached sooner for this user due to the lack of linguistic behaviour viable for coding). The accounts of users within this high scoring group are therefore suggestive of narcissistic traits. However, despite being grouped based on the word frequency-derived metric of grandiose narcissism, these traits are described in terms of themes more compatible with vulnerable narcissism.

The most conspicuous feature of users in the low narcissism group was their lack of any emphasised themes, being instead characterised by a consistent set of themes that occurred with a lower frequency (i.e. data saturation was reached sooner for these users due to the lack of linguistic behaviour viable for coding). The notable exceptions were users **H** and **J**. The former was characterised by a much greater tendency for emphasising disengagement and avoidance and narcissistic beliefs, whereas the latter was similar to users in the high grandiose narcissism group and

displayed a markedly greater disposition towards disengagement and avoidance, hostility towards others, and narcissistic signalling behaviour. The accounts of both users are therefore as legitimately suggestive of stronger narcissistic traits as are the accounts described above for the high grandiose narcissism group, although similarly reflect narcissistic traits more akin to vulnerability than grandiosity.

In summary, the qualitative differences between the high and low grandiose narcissism groups are somewhat consistent with their word frequency-derived categorization. Users in the high group on the whole displayed an appropriate set of traits consistent with a narcissistic persona. The higher code frequency and corresponding higher threshold for data saturation in this group also indicates the presence of more viable narcissistic linguistic behaviours, which is similarly indicative of higher trait narcissism. In contrast, users in the low narcissism group displayed an overall less clearly delineated set of narcissistic traits with a lower code frequency. However, not all users in each group were consistent with their word frequency-derived categorization, and narcissistic traits emphasised irrespective of group type were more characteristic of narcissistic vulnerability than grandiosity.

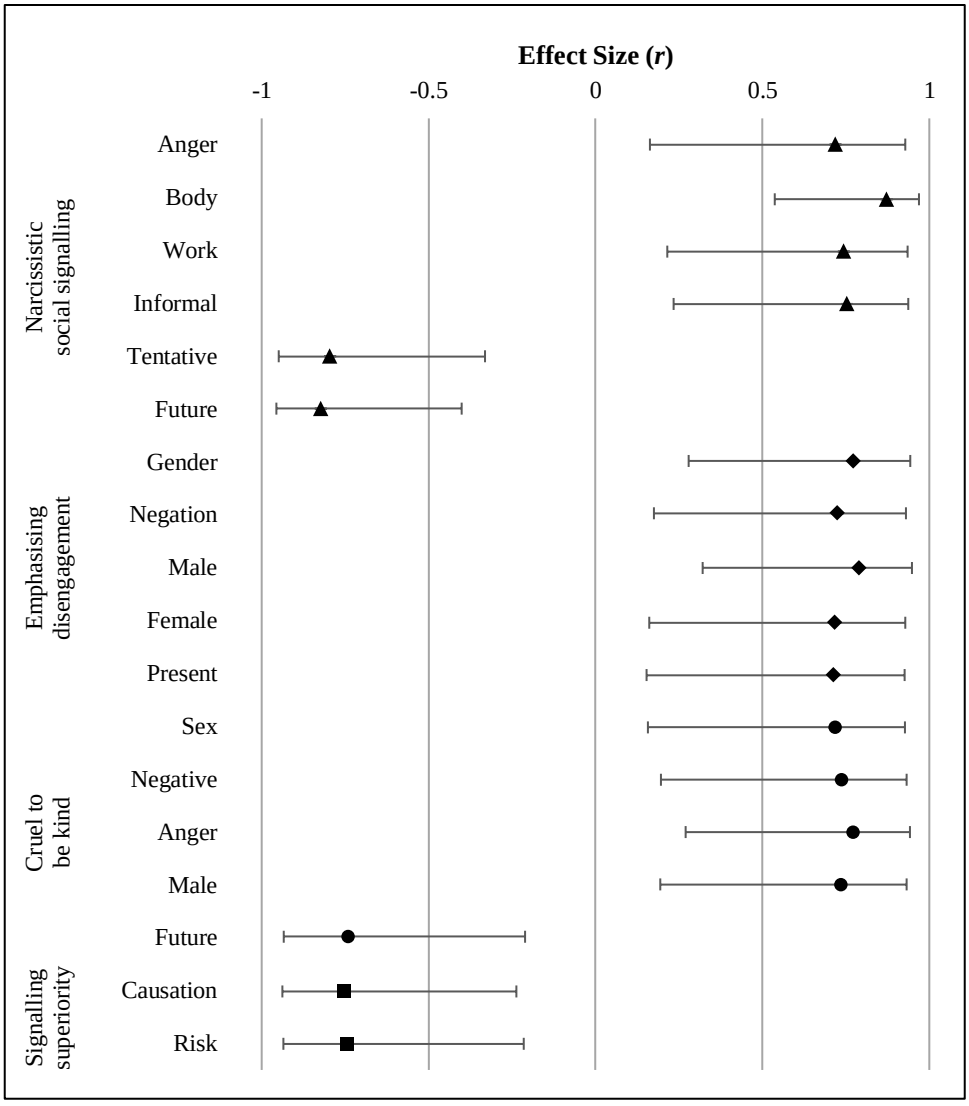
Relationship between qualitative narcissistic traits and word frequency metrics

Zero-order correlations with 95% CIs between code frequencies for identified themes and word frequency metrics are presented in **Figure 2**. To control for inflated effect sizes due to the low sample size used in the analysis, only statistically significant correlations with a coefficient greater than 0.7 are reported (Dancey & Reidy, 2007).

Narcissistic social signalling was strongly positively correlated with anger (r = .719, p = .019, 95% CI [.163, .928]), body-related language (r = .872, p = .001, 95% CI [.537; .969]), work-related language (r = .744, p = .014, 95% CI [.215; .935]) and informal language (r = .753, p = .012, 95% CI [.234; .937]), and strongly negatively correlated with tentative language (r = -.795, p = .006, 95% CI [-.949; -.331]) and future-focused language (r = -.823, p = .003, 95% CI [-.956; -.401]). No statistically significant correlations with $r > 0.7$ were observed for narcissistic beliefs. No statistically significant correlations with $r > 0.7$ were observed for passive hostility as

a subtheme of guarded vulnerability. However, the subtheme emphasising disengagement and avoidance was strongly positively correlated with use of gendered pronouns (r = .773, p = .009, 95% CI [.279; .943]), negations (r = .725, p = .018, 95% CI [.175; .930]), male-related language (r = .791, p = .006, 95% CI [.321; .948]), female-related language (r = .718, p = .019, 95% CI [.161; .928]), and present-focused language (r = .714, p = .020, 95% CI [.153; .926]).

Figure 2. Zero-order correlations with 95% CIs between code frequencies for identified sub-themes and word frequency metrics.



Cruel to be kind as a subtheme of assertive grandiosity was strongly positively correlated with sex and sexuality (r = .716, p = .020, 95% CI [.157; .927]), negative

emotional language (r = .735, p = .015, 95% CI [.196; .932]), anger (r = .769, p = .009, 95% CI [.270; .942]), male-related language (r = .734, p = .016, 95% CI [.194; .932]), and strongly negatively correlated with future-focused language (r = -.742, p = .014, 95% CI [-.934; -.211]). Signalling superiority was strongly negatively related with language related to causation (r = -.754, p = .012, 95% CI [-.938; -.237])) and risk related language (r = -.744, p = .014, 95% CI [-.935; -.215]). No statistically significant correlations with r > 0.7 were observed for righteous hostility. A description of each theme as a novel word frequency-derived metric in terms of these associations is presented in **Table 2**.

Table 2. Associations between each theme/sub-theme and its associated word frequency categories.

	Associated word frequency categories
Narcissistic social signalling	
-	Anger (+), Body (+), Work (+), Informal (+), Tentative (-), Future (-)
Guarded vulnerability	
Emphasising avoidance and disengagement	Gender (+), Negation (+), Male (+), Female (+), Present (+)
Assertive grandiosity	
Cruel to be kind	Sex (+), Negative Affect (+), Anger (+), Male (+), Future (-)
Signalling superiority	Causation (-), Risk (-)

Sub-themes are denoted in italics. Positive/negative associations are denoted by (+)/(-) respectively.

Summary of results

The most frequent contributors from the 'r/relationships' Reddit community were sorted by a word frequency-derived metric of grandiose narcissism and divided into groups containing the five highest and lowest scoring users. This sample of users were anonymized and randomly sorted. Thematic analysis of these users identified seven themes and sub-themes representing both general and subtype specific differences in narcissistic behaviour and beliefs, including some relevant to narcissistic aggression. Whilst recognizably narcissistic in their features, the themes identified were all clearly influenced by a meta-theme representing motivation to adhere to community norms and only engage in narcissistic behaviours in contexts where it could be construed as appropriate. Themes were therefore by no means representative of the classic comprehensive set of narcissistic traits that define psychometric operationalization of narcissism. Instead, narcissistic linguistic behaviours as observed in the naturalistic dataset used by the present study emphasised a more specific set of features, delineating trait-based accounts of narcissism from an account of narcissism situated in dynamic social behaviour.

The word frequency-derived metric of grandiose narcissism identified by previous research and generated using the same naturalistic dataset was largely in agreement with this qualitative assessment, with users in the high grandiose narcissism displaying markedly more narcissistic features compared to users in the low scoring group. However, interestingly, these narcissistic features were characterised more by narcissistic vulnerability than grandiosity (see below for further discussion). This highlights both the potential usefulness of word frequency-derived metrics, and the validity of the use of this metric as a means of selecting the final sample of users featured in the present study.

Furthermore, this validates the exploration and development of other word frequency-derived metrics driven by the narcissistic features identified by the thematic analysis. The synthesis of this quantitative and qualitative data in the present study produced three clearly defined word frequency metrics of different narcissistic traits, with a fourth less clearly defined metric also potentially being identified. These metrics

correspond to narcissistic social signalling behaviours, vulnerable avoidance behaviours, callous grandiose behaviours, and superior grandiose behaviours respectively. Despite some methodological limitations with this synthesis, the coherency of the metrics identified is encouraging.

5.5. Discussion

Qualitative investigation of narcissistic behaviour in an online community

As described by the theoretical framework of narcissism used throughout this thesis, narcissism is characterised by a suite of cognitive and behavioural strategies that serve to maintain narcissistic self-perceptions, which due to their inherent incompatibility with reality are fragile and easily challenged (Raskin et al., 1991). These strategies therefore also serve to defend these narcissistic self-perceptions from potential challenges. Defensive strategies can be both reactive and proactive - in the case of the former, aggressive hyperactivating or deactivating strategies provide a fast and effective solution, whereas in the case of the latter, narcissistic social engineering strategies are emphasised. Social engineering can feature direct or indirect signalling of attributes evidencing narcissistic perceptions (i.e. signalling attributes personal to oneself or relative to others), or behaviours intended to influence the social context to become more facilitative to narcissistic perceptions, such as encouraging narcissistic beliefs and behaviours, selectively curating a facilitative social context, or aggressively asserting narcissistic beliefs (Grapsas et al., 2020). Subtype differences manifest as dispositions towards particular strategies, with grandiose narcissism typically being more assertive and less concerned with the actions of others, and vulnerable narcissism being more avoidant, hostile and neurotic (Miller et al., 2011).

However, the results of the present study expand on this framework and stress the relevance of communal narcissism as a subtype difference, emphasising that the relationship between narcissistic behaviours and beliefs and the social context in which they occur is not one-way, but rather the social context itself imposes limitations that influence the manifestation of narcissistic traits. Community factors manifested clearly as an overarching meta-theme ‘Conforming behaviour to

community influence'. The factors described by this meta-theme temper the manifestation of typically grandiose narcissistic signalling, whilst also providing a means of legitimising narcissistic beliefs by embedding them in superficially prosocial qualities, and providing a socially acceptable means with which to express these beliefs (Meade & Castle, 2021). Social media provides an effective platform for curating a persona ideally suited to successfully navigate these community factors for the benefit of the narcissistic individual (Kristinsdottir et al., 2021; Meade & Castle, 2021), with simple and highly rewarding metrics of community approval mechanisms to incentivize narcissistic behaviours (Tuchband, 2018). Indeed, there is a straightforward association between narcissism and social media activity (Buffardi & Campbell, 2008).

The influence of this meta-theme was clearly visible in the set of themes identified by the thematic analysis. Specifically, the themes identified were 'narcissistic social signalling' and 'narcissistic beliefs', which represented general (i.e. non-subtype specific) narcissistic linguistic behaviours maintaining narcissistic self-perceptions. Social signalling behaviours emphasised indirect signalling through the shaming of others, implicitly emphasising positive personal qualities relative to the negative qualities of others, and maintaining a social context of lower status individuals. The expression of these beliefs may represent an assertive narcissistic tendency to perceive social interactions as competitive situations where one is established as being either higher or lower in status, and thereby necessitating the use of 'dominance' strategies focused on lowering the status of others (Zeigler-Hill et al., 2021). Social media platforms may incentivize this antagonistic perception of social interactions by emphasising the need to compete with other users for attention (Balcerowska & Sawicki, 2022).

Similarly, beliefs emphasised self-centred approaches to navigating relationships, ignoring the negativity of others and condoning the use of manipulation, representing both a genuine expression of narcissistic beliefs and an effort to normalise these beliefs. However, importantly, the expression of both signalling strategies and beliefs was clearly framed as prosocial behaviour, engaging in shaming behaviour as a

community activity and to enforce community virtues, or expressing narcissistic beliefs to help an other within the community solve a problem. This type of behaviour is highly characteristic of communal narcissism (Gebauer et al., 2012; Gebauer & Sedikides, 2018), with otherwise undesirable narcissistic beliefs obfuscated by desirable prosocial behaviours implicitly signalling these narcissistic beliefs. This aspect of communal narcissism has been described in the literature as 'pathological altruism' (Kaufman & Jauk, 2020; Oakley et al., 2011), which broadly refers to acts that are ultimately antisocial in nature and/or outcome despite being superficially prosocial. Pathological altruism is driven by a desire to avoid rejection and criticism from a community, and gain community approval and status, and as a result is motivated primarily by self-interested goals more than genuinely prosocial motivations (Kaufman & Jauk, 2020). The observed intersection between narcissistic beliefs, behaviours and community influences may therefore be described as an instance of pathological altruism. Indeed, despite expressing prosocial intentions, research suggests communal narcissists do not necessarily manifest these intentions in behaviour, whilst measures of physiological reactivity suggest they continue to display classic features of narcissistic hypersensitivity that underly the darker side of narcissistic cognition (Yang et al., 2018).

Subtype differences manifested as four sub-themes split between two themes. Vulnerable narcissistic linguistic behaviours identified by the thematic analysis included 'passive hostility' and 'emphasising disengagement and avoidance'. The former emphasised distrust and pessimism, and stand-offish attitudes towards others veiled in apparently passive and non-confrontational language. The latter encouraged avoiding confrontation and avoiding investing resources in challenging situations where leaving or ignoring the cause is viable. These cynical attitudes are associated with antagonistic narcissistic beliefs, and may reflect the tendency to view social interactions as having a competitive and potentially risky aspect that belies their superficially prosocial function (Szymczak et al., 2020).

Grandiose narcissistic linguistic behaviours identified included 'cruel to be kind' and 'signalling superiority', the former involving callous behaviours with the superficial

intent of unambiguously directing and supervising the actions of others, and the latter involving directly signalling perceived positive qualities and achievements. Whilst signalling of superiority can be interpreted as a straightforward manifestation of expressing grandiose beliefs, linguistic behaviours that are callous in nature are less intuitively described by communal narcissistic motivations. Pathological altruism may therefore be useful again here as a concept, with cruel or harmful behaviour performed with apparently helpful intentions being a consequence of overcommitment to prosocial behaviours. This overcommitment stems from the desire to maximize community approval and acceptance, but becomes paradoxically antisocial due to the motivations for this desire being ultimately self-interested and driven by fear of rejection (Kaufman & Jauk, 2020; Oakley et al., 2011). Interestingly, this description highlights that even if grandiose prosocial behaviours are inherently self-interested and ultimately harmful, they may be interpreted as genuinely prosocial by the narcissistic individual. Narcissistic extraversion is associated with trusting attitudes towards others (Szymczak et al., 2020), and in online community contexts this trust that others will interpret superficially callous behaviours as well-intentioned may be an important feature of grandiose narcissistic cognition. Narcissistic individuals should therefore not be associated as inherently bad faith actors in online communities, with self-centred and manipulative intentions occurring alongside misguided but genuine intentions to develop a role within the community and achieve acceptance (Shanahan et al., 2019).

The vulnerable and grandiose themes identified are therefore coherent with vulnerable narcissistic dispositions towards neuroticism, hostility and avoidance-oriented behaviours, and grandiose narcissistic dispositions towards assertive and superior behaviours, but are similarly selective in their manifestation due to community influences. For example, conspicuously lacking are overtly assertive grandiose behaviours or overtly reactive vulnerable behaviours, with these behaviours always limited by the necessity of community conduct. Similarly lacking are expressions of overtly self-focused and grandiose beliefs independent of prosocial justification. That is to say, any expression of grandiose beliefs was controlled and dictated by community norms that require beliefs to be justified and expressed only in

an appropriate context. This is coherent with the finding reported elsewhere that classically agentic narcissistic linguistic behaviour is viewed more harshly by social media users than other forms of narcissism (Joyce et al., 2019), with communal narcissists being more popular by virtue of making some effort to reciprocate the benefits of community membership by engaging in prosocial behaviours (Rentzsch & Gebauer, 2019). Community influences therefore introduce prosocial concessions to the manifestation of otherwise highly individualistic grandiose beliefs.

Conversely, conspicuously present are vulnerable narcissistic behaviours and beliefs that express neuroticism, hostility and avoidance, despite the users expressing them apparently voluntarily participating in a superficially prosocial activity that involves engaging with others. This contradiction may represent a disconnect between the expression of vulnerable narcissistic beliefs in 'real life' contexts versus the context of an online community. For example, whilst an individual in an online community may condone and relate to vulnerable narcissistic behaviour in terms of their 'real life' experiences, they may not exhibit this in their behaviour in online communities. Indeed, as described above, the fusion of narcissistic beliefs with superficially prosocial justifications may result in a disconnect between these beliefs and actual behaviour (Yang et al., 2018). Therefore, whilst community influences control and limit the expression of grandiose narcissistic beliefs, they may facilitate the expression of some vulnerable narcissistic beliefs. This facilitation is likely due to the relatively safe space a structured and moderated online community offers, providing layers of social protocol protecting and justifying expression of beliefs. For individuals driven to pursue highly agentic forms of narcissistic behaviour, this may prove frustrating and act as a deterrent to participation. However, exploiting the safe and structured nature of online communities may be enticing for individuals who are motivated pursue communal forms of narcissistic self-enhancement (Balcerowska & Sawicki, 2022). The apparent disconnect between avoidant beliefs and avoidant behaviour may also be due to the lack of a meaningful way to disengage without removing oneself from the community altogether. For example, as the community is structured as an open forum, and it is not possible to withdraw and selectively engage with only a subset of individuals. This renders engagement an all-or-nothing activity,

that prevents the type of avoidant behaviour emphasised by vulnerable narcissism, wherein an individual does not want to exile themselves wholesale but rather control their social context.

Grandiose and vulnerable linguistic behaviours therefore remain distinct, but become blurred by virtue of communal influences rendering grandiose narcissism slightly more cautious and vulnerable narcissism slightly more bold. Irrespective of their distinct features, this shared presence of both subtype traits is consistent with research suggesting social media use is similarly associated with both grandiose and vulnerable subtypes, with social media platforms facilitating shared underlying motivations and anxieties (Brailovskaia et al., 2020). More generally, qualitative accounts of narcissism have noted the simultaneous occurrence of both vulnerable and grandiose traits (Day et al., 2020). This may result in both subtypes becoming harder to meaningfully distinguish from one another at a trait level, and emphasises the potential utility of a more general communal narcissism subtype reflecting elements of both.

Finally, narcissistic aggression manifested as the theme ‘righteous hostility’, which consisted of linguistic behaviours that were directly critical, explicitly intended to shock, or explicitly insulting or verbally aggressive. These behaviours naturally represent the only means a user in a large online community has to behave aggressively towards others. This is not to suggest that virtual aggressive behaviour (or ‘cyberaggression’) is not diverse, but the existing research literature examines cyberaggression almost exclusively in terms of repeated and individually targeted harmful or humiliating actions. These forms of aggression are more akin to cyberbullying, and have less in common with the tacitly approved forms of cyberaggression investigated in the present study (Grigg, 2010; Pyżalski, 2012). Instances of ‘communal narcissistic cyberaggression’ may feature some of the same behaviours, but are a relatively non-specific act that is part of a larger behavioural system pursuing community approval. The results of the present study therefore offer insight into this phenomenon that is largely unrepresented in the literature. In the context of the ‘r/relationships’ community, the explicit focus on providing advice and

the implicit focus on passing judgement on the behaviour and beliefs of others provides a socially acceptable outlet for aggression through the 'invited' criticism of others. It is important to note that whilst this linguistic behaviour is taken to represent manifestations of narcissistic aggression, narcissistic aggression here is simply defined as any aggressive behaviour occurring alongside narcissistic linguistic behaviours. Whilst this may be a simplistic approach, there is no definitive way of categorising the hostile and aggressive behaviour observed as uniquely narcissistic. These behaviours are naturally not exclusive to narcissism, and are therefore inherently more difficult to code precisely compared to linguistic behaviours that are arguably more clearly narcissistic. Instead, the results of the thematic analysis highlight how aggression manifests in online communities, including narcissistic aggression, with the motivations potentially driving this aggression in narcissistic individuals not superficially apparent, or only indirectly apparent after-the-fact with relation to narcissistic traits identified by the thematic analysis.

In summary, narcissistic behaviour in online communities is dynamic and clearly shaped by community influences. This represents a concession between narcissistic motivations and the need to leverage community approval to achieve these motivations. However, a dimension of this behaviour that is inaccessible in the present study is what the outcome of this behaviour is over time, and how narcissistic individuals adapt their behaviour to suit the changing nature of their relationship with the community. Research suggests narcissistic linguistic behaviour is relatively stable over time and between contexts (Fast & Funder, 2008), and can signal charismatic traits that are attractive to potential followers (Brunzel, 2021). However, it seems likely that the pathologically altruistic behaviour associated with communal narcissism is unlikely to be a viable long-term strategy for achieving community approval. Indeed, research investigating narcissism in corporate communities suggests superficially prosocial but ultimately self-interested behaviours are not sufficient to maintain community approval, with relationships deteriorating over time (Braun, 2017). This may be particularly true for social media communities, which may exaggerate narcissistic behaviours and promote the adjacent development of

neurotic or depressive cognition (Alimohammadi et al., 2021), accelerating this deterioration.

Critique of the word frequency-derived grandiose narcissism metric

Previous research identified a word frequency-derived metric of grandiose narcissism, that was calculated based on text analysis of the dataset featured in the present study and used to select the final sample of users included in the thematic analysis. The accuracy of this grouping in terms of its compatibility with a qualitatively-derived assessment of narcissism is described below. Statistical assessments of its utility as a metric were not encouraging, with Cronbach's alpha suggesting very low internal reliability. It could be that the community from which the sample was taken influenced the type of language used, and therefore confounded the relationship between word frequency categories used in the metric. However, given the relatively unrestricted nature of the community, there is little reason to suggest a strong community influence of this kind. However, whilst the study that previously identified the metric (Holtzman et al., 2019) was based partially on data from Facebook (i.e. another large online community presumably governed by similar community influences), it was also partially derived from essays written by participants, which may have featured use of sufficiently different language to render the metric less appropriate to the dataset featured in the present study (Schwartz et al., 2013).

Regardless, it was not assumed that this metric would necessarily be a reliable and valid metric of grandiose narcissism, and it was nevertheless useful in the present study given the lack of any other means of impartially assessing narcissistic traits in the sample. Furthermore, one of the objectives of the present study was validating this metric and making an attempt to replicate the findings of previous research (which has been both a goal of much of the work documented earlier in this thesis, and is of

course an important responsibility of psychological research more generally). As a result, the low reliability of this metric of grandiose narcissism is presented here as a result rather than a limitation.

Evaluating qualitative and word frequency-derived accounts of grandiose narcissism

Despite concerns regarding the reliability of the word frequency-derived metric of grandiose narcissism, a qualitative assessment of its assessment suggests it was largely successful as a means of sorting users into groups exhibiting high and low trait narcissism. Evaluating each user in the final sample revealed clear differences between groups with only some variability present. Variability in the individuals identified by this metric would be expected, given both the untested nature of the metric, and the possibility of narcissistic linguistic behaviours occurring as a result of general/subtype-specific narcissistic traits other than grandiose narcissism. Importantly, these differences are not apparently arbitrary, but rather represent coherent differences that one might expect to observe between groups, with users in the high narcissism group exhibiting a richer set of associated themes (indicating a higher threshold for data saturation), and a markedly higher disposition for shaming, criticism, cynical views and emphasising disengagement from relationships.

This word frequency-derived metric therefore seems to have been able to select two groups with meaningfully different traits that resemble differences in trait narcissism. This is coherent with findings from the broader individual differences text analysis literature, which have reported the successful predictive use of word frequency metrics for personality traits (Kern et al., 2014; Schwartz et al., 2013; Yarkoni, 2010). However, it should be noted that the choice to describe these between-groups differences as narcissism rather than grandiose narcissism is deliberate. The qualitative assessment of users revealed no notable emphasis (or notable lack thereof) for themes relating to grandiose narcissistic linguistic behaviour, such as callous or strongly directive language, or direct signalling of superiority. Instead, themes

emphasised more closely resembled vulnerable narcissism, being characterised by a combination of linguistic behaviours featuring general narcissistic signalling strategies and vulnerable narcissistic hostile perceptions. As discussed above, it is possible that manifestations of grandiose and vulnerable traits are harder to disentangle in the context of the naturalistic dataset featured in the present study, where the former may be more cautious than expected and the latter more assertive, potentially explaining this apparent crossover of subtype traits.

Regardless of the explanation, the fact that it is possible to make this observation speaks to the strengths of thematic analysis (or a similar qualitative analysis more generally) as a method. Without the means to perform a qualitative assessment of the users in the sample, it would have been entirely reasonable to conclude that the word frequency-derived metric represents differences in grandiose narcissistic traits, and potentially use this conclusion to interpret the relationship between other quantitative traits and grandiose narcissism. However, with the aid of the rich qualitative evaluation of the individuals described by the metric, this conclusion is shown to be incorrect. This highlights that the way in which traits manifest on a personal level (i.e. as measured by psychometrics) may not necessarily be representative of how traits manifest in behaviour. The extent to which this challenges existing text analaysis research investigating manifestations of narcissism (frequently as a 'dark triad' trait) is uncertain but interesting to consider, with other studies of large online datasets taking more traditional self-report approaches in which word frequency patterns are derived from psychometric scores (Bogolyubova et al., 2018; Preotiuc-Pietro et al., 2016; Sumner et al., 2012). Whilst these studies appear robust and coherent in their results, the results of the synthesized analysis described here identify potential blindspots that may be missed without qualitative insight.

Use of code frequency as a mixed methods metric

Use of code frequency as a metric features commonly in some other traditionally mixed methods qualitative approaches such as content analysis (Roller & Lavrakas, 2015; for examples of this approach applied to narcissism see Paramboukis et al.,

2016; Utama et al., 2020), and can be used as a descriptive tool in thematic analysis approaches (e.g. Grayman-Simpson, 2009). Here, however, the qualitative analysis is conducted independently of any reference to code frequency as a thematic analysis, in order to generate an independent qualitative dataset. Code frequencies are then derived from this dataset to be used in an independent synthesis analysis, but this code frequency data should be considered a secondary product of the thematic analysis and not its primary output. This distinction is important to clarify the difference between the thematic analysis method used in the present study and a content analysis method. Whilst code frequency metrics are often used for descriptive purposes or to facilitate qualitative analysis, their more formal use as a metric in quantitative analyses (as in the present study) is also possible (Armborst, 2017). However regardless one must be careful to consider what code frequencies are assumed to represent, acknowledging the context of the qualitative data from which they are derived (Roller & Lavrakas, 2015). For example, code frequency metrics are occasionally defined not as the overall quantity of codes, but as the quantity of individuals exhibiting a given code. This is typically done in order to avoid the misleading effect of contextual factors that might exaggerate the frequency of certain codes (e.g. mentions of violence in a video game community where fictionalised violence is common). In the present study, as narcissistic behaviours and beliefs are expected to be present to at least some extent in most individuals (as with any non-clinical psychological trait), defining frequency as 'individuals per code' was less appropriate than 'codes per individual', as the latter is better suited to investigating individual differences. To avoid the misleading effect of defining code frequency this way, careful attention is paid to contextual factors present in the dataset (e.g. community norms, the topic being discussed), and this synthesis itself with a text analysis metric measuring (at least partially) the same phenomenon allows for further examination and validation of code frequency-derived metrics.

Qualitatively driven development of novel word frequency-derived metrics of narcissism

An exploratory analysis approach was taken to synthesise the results of the text analysis and the thematic analysis, in order to use the qualitative data generated by the latter to drive the identification of novel word frequency metrics from within the data generated by the former. Word frequency metrics were identified for 'narcissistic social signalling', 'emphasising disengagement/avoidance', 'cruel to be kind', and 'signalling superiority'. The specific word frequency associations characterizing each metric were consistent with the narcissistic traits described by the metric, with anger, negative emotion, impulsivity and avoidance featuring where appropriate.

The present study is not only the first study of its kind to adopt this mixed methods approach to studying narcissism, but also the first study to apply text analysis methods to studying narcissistic subtype differences. Indeed, as discussed earlier in the thesis (see **Chapter 1**), there is a tendency for research to take a 'broad strokes' approach to narcissism, and feature a description of narcissism that would be more appropriately described as the grandiose subtype. This is clearly evident in the text analysis literature. As a result, a clear comparison with the results of similar research is difficult to achieve. However, insofar as the generalised conception of narcissism used in other research can be considered equivalent to the general overview of the results of the present study, then the results are coherent. Holtzman et al. (2010) report a similar association with anger and swear words (the latter reflected by 'informal' word frequency in the results), whereas Ireland & Mehl (2014) report a similar association with anger, negative emotion, and swear words. Holtzman et al. (2019), who more formally identify their operationalisation of narcissism as grandiose, and whose finding informed the word frequency metric used in the present study, report a less similar set of associations, but find similar association with swear words and a similar negative association with tentative words. Importantly, whilst occasionally identifying different associations, the associations identified in the literature do not conflict with those identified in the present study.

Some word frequency associations were most likely a product of the dataset rather than a product of narcissistic traits. For example, the increased frequency of sex, body

and work-related words observed for ‘narcissistic social signalling’ reflects relationship advice frequently referencing physical interactions and lifestyle activities, although there is an established association between narcissism and sexual language (Holtzman et al., 2010, 2019; Ireland & Mehl, 2014). Similarly, ‘emphasising disengagement’ and ‘cruel to be kind’ were associated with increased frequency of gender-related words and pronouns (i.e. explicitly gender related, such as “he”, “she”, “boy”, “mother”). This may be a product of advice often necessarily involving reference to another individual involved in a relationship with the recipient, and therefore involving some purely grammatical reference to the gender of that individual.

No word frequency-derived metrics were observed for the other themes identified by the thematic analysis. As discussed above, it is difficult to say for certain whether other word frequency metrics should be expected, as the few relevant studies in the literature are limited to examining grandiose forms of narcissism (either implicitly or explicitly). Studies examining the association between word frequency and ‘Big Five’ traits have reported an association between neuroticism and negative emotion words (Kern et al., 2014; Schwartz et al., 2013; Yarkoni, 2010), which may suggest an association between this category and themes relating to vulnerable narcissism should have been expected (Zajenkowski & Szymaniak, 2021). It is likely that this lack of other effects is at least partially a result of the relatively conservative threshold for effect size necessary for inclusion in the results. This conservative approach was a necessary consequence of the small sample size featured in the synthesis, which itself necessarily follows from the restricted sample size required by the thematic analysis. As a result, it is possible that the metrics identified above reflect only the strongest associations between themes and word-frequency.

With regards to the strength of associations, it is important to note that the effect size confidence intervals for the observed associations between themes and word frequencies were large. This is anticipated due to the low sample size, and suggests the true effect size of the otherwise strong effects observed may be much weaker in

the wider population. Population effect sizes for text analysis metrics tend to be quite low (Kern et al., 2014), likely due to linguistic behaviours being highly dynamic in practice (Schwartz et al., 2013; Holtzman et al., 2019), and so it is reasonable to conclude that the true effect size in this dataset is likely closer to the lower end of the confidence intervals observed for the sample in the present study. This is not necessarily problematic in terms of drawing conclusions regarding the statistical significance of the effects observed, if the true effect size is expected to be lower globally for text analysis metrics, but nevertheless it is necessary to acknowledge this effect of statistical distortion. Regardless, as the sample of users featured in the analysis was structured to be as heterogeneous as possible, it is encouraging that effect sizes observed were large in spite of this, as it is likely this heterogeneity acted as a limiting factor. Finally, given the large number of variables included in the synthesis analysis, it is also important to consider the potential issue of multiple comparisons. However, given the relatively large number of observed statistically significant associations with effect sizes above this threshold, and given that the pattern of associations observed were theoretically coherent and compatible with one another, concerns regarding the effect of multiple comparisons seem unfounded.

Conclusions

The present study represents a proof of concept, demonstrating that large naturalistic datasets are both easily accessible via the use of web scraping and capable of generating meaningful data. With regards to the investigation of narcissism and narcissistic aggression, qualitative methods can produce insights from this data that are not evident with quantitative methods or with non-naturalistic datasets. Specifically, the results of the thematic analysis suggest that traits do not manifest archetypally but rather manifest selectively based on the influence of community norms, contrary to what may be predicted based on psychometric lab-based accounts of narcissism. These communal manifestations of narcissism are less strongly delineated by subtype differences and emphasise a distinct set of narcissistic behaviours, highlighting the utility of communal narcissism as a concept and an independent subtype trait. The present study also demonstrated the viability of a

synthesis of qualitative and quantitative word frequency data, generating several new word frequency-derived metrics relating to narcissistic traits. These may provide useful tools for individuals exploring narcissism in similar naturalistic datasets using text analysis approaches. Whilst a previously identified word frequency-derived metric of grandiose narcissism was not successfully validated, it nevertheless proved useful and largely consistent with qualitatively assessed narcissistic traits. Subsequent research could provide useful further validation of both this metric and the novel metrics identified in the present study.

References

Alimohammadi, A., Mortazavi, S. S., Atef-Vahid, M.-K., & Shahbazi, N. (2021). Psychoanalytic Explanation of Key Personality Traits of Instagram Micro-celebrities: A Qualitative Study. *Iranian Journal of Psychiatry and Behavioral Sciences*, *15*(2), Article 2. https://doi.org/10.5812/ijpbs.108666

Armborst, A. (2017). Thematic Proximity in Content Analysis. *SAGE Open*, *7*(2), 2158244017707797. https://doi.org/10.1177/2158244017707797

Balcerowska, J. M., & Sawicki, A. J. (2022). Which aspects of narcissism are related to Social Networking Sites addiction? The role of self-enhancement and self-protection. *Personality and Individual Differences*, *190*, 111530. https://doi.org/10.1016/j.paid.2022.111530

Bogolyubova, O., Panicheva, P., Tikhonov, R., Ivanov, V., & Ledovaya, Y. (2018). Dark personalities on Facebook: Harmful online behaviors and language. *Computers in Human Behavior*, *78*, 151–159. https://doi.org/10.1016/j.chb.2017.09.032

Brailovskaia, J., Rohmann, E., Bierhoff, H.-W., & Margraf, J. (2020). The anxious addictive narcissist: The relationship between grandiose and vulnerable narcissism, anxiety symptoms and Facebook Addiction. *PLOS ONE*, *15*(11), e0241632. https://doi.org/10.1371/journal.pone.0241632

Braun, S. (2017). Leader Narcissism and Outcomes in Organizations: A Review at Multiple Levels of Analysis and Implications for Future Research. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/article/10.3389/fpsyg.2017.00773

Braun, V., & Clarke, V. (2021). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health*, *13*(2), 201–216. https://doi.org/10.1080/2159676X.2019.1704846

Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic Analysis. In P. Liamputtong (Ed.), *Handbook of Research Methods in Health Social Sciences* (pp. 843–860). Springer. https://doi.org/10.1007/978-981-10-5251-4_103

Brunzel, J. (2021). Overconfidence and narcissism among the upper echelons: A systematic literature review. *Management Review Quarterly*, *71*(3), 585–623. https://doi.org/10.1007/s11301-020-00194-6

Buffardi, L. E., & Campbell, W. K. (2008). Narcissism and Social Networking Web Sites. *Personality and Social Psychology Bulletin*, *34*(10), 1303–1314. https://doi.org/10.1177/0146167208320061

Dancey, C. P., & Reidy, J. (2007). *Statistics Without Maths for Psychology*. Pearson Education.

Day, N. J. S., Townsend, M. L., & Grenyer, B. F. S. (2020). Living with pathological narcissism: A qualitative study. *Borderline Personality Disorder and Emotion Dysregulation*, *7*(1), 19. https://doi.org/10.1186/s40479-020-00132-8

Denissen, J. J. A., van Aken, M. A. G., Penke, L., & Wood, D. (2013). Self-Regulation Underlies Temperament and Personality: An Integrative Developmental Framework. *Child Development Perspectives*, *7*(4), 255–260. https://doi.org/10.1111/cdep.12050

Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology*, *94*(2), 334–346. https://doi.org/10.1037/0022-3514.94.2.334

Gebauer, J. E., & Sedikides, C. (2018). Communal Narcissism: Theoretical and Empirical Support. In A. D. Hermann, A. B. Brunell, & J. D. Foster (Eds.), *Handbook of Trait Narcissism: Key Advances, Research Methods, and*

Controversies (pp. 69–77). Springer International Publishing. https://doi.org/10.1007/978-3-319-92171-6_7

Gebauer, J. E., Sedikides, C., Verplanken, B., & Maio, G. R. (2012). Communal narcissism. *Journal of Personality and Social Psychology*, *103*(5), 854–878. https://doi.org/10.1037/a0029629

Grapsas, S., Brummelman, E., Back, M. D., & Denissen, J. J. A. (2020). The "Why" and "How" of Narcissism: A Process Model of Narcissistic Status Pursuit. *Perspectives on Psychological Science*, *15*(1), 150–172. https://doi.org/10.1177/1745691619873350

Grayman-Simpson, N. (2009). We Who Are Dark...The Black Community According To Black Americans. *Journal of Black Psychology*.

Grigg, D. W. (2010). Cyber-Aggression: Definition and Concept of Cyberbullying. *Journal of Psychologists and Counsellors in Schools*, *20*(2), 143–156. https://doi.org/10.1375/ajgc.20.2.143

Holtzman, N. S., Tackman, A. M., Carey, A. L., Brucks, M. S., Küfner, A. C. P., Deters, F. G., Back, M. D., Donnellan, M. B., Pennebaker, J. W., Sherman, R. A., & Mehl, M. R. (2019). Linguistic Markers of Grandiose Narcissism: A LIWC Analysis of 15 Samples. *Journal of Language and Social Psychology*, *38*(5–6), 773–786. https://doi.org/10.1177/0261927X19871084

Holtzman, N. S., Vazire, S., & Mehl, M. R. (2010). Sounds like a Narcissist: Behavioral Manifestations of Narcissism in Everyday Life. *Journal of Research in Personality*, *44*(4), 478–484. https://doi.org/10.1016/j.jrp.2010.06.001

Ireland, M. E., & Mehl, M. R. (2014). Natural language use as a marker of personality. In *The Oxford handbook of language and social psychology* (pp. 201–218). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199838639.001.0001

Jauk, E., Weigle, E., Lehmann, K., Benedek, M., & Neubauer, A. C. (2017). The Relationship between Grandiose and Vulnerable (Hypersensitive) Narcissism. *Frontiers in Psychology*, *8*, 1600. https://doi.org/10.3389/fpsyg.2017.01600

Joyce, L., Chrysosferidis, J., Mousavi, S., Barry, C., & Benson, C. (2019). Perceptions of Agentic and Communal Narcissism on Facebook.

Cyberpsychology, Behavior, and Social Networking. https://doi.org/10.1089/cyber.2019.0135

Kaufman, S. B., & Jauk, E. (2020). Healthy Selfishness and Pathological Altruism: Measuring Two Paradoxical Forms of Selfishness. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/article/10.3389/fpsyg.2020.01006

Kennedy, B., Ashokkumar, A., Boyd, R., & Dehghani, M. (2021). *Text Analysis for Psychology: Methods, Principles, and Practices*. https://doi.org/10.31234/osf.io/h2b8t

Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Dziurzynski, L., Ungar, L. H., Stillwell, D. J., Kosinski, M., Ramones, S. M., & Seligman, M. E. P. (2014). The Online Social Self: An Open Vocabulary Approach to Personality. *Assessment*, *21*(2), 158–169. https://doi.org/10.1177/1073191113514104

Kristinsdottir, K. H., Gylfason, H. F., & Sigurvinsdottir, R. (2021). Narcissism and Social Media: The Role of Communal Narcissism. *International Journal of Environmental Research and Public Health*, *18*(19), 10106. https://doi.org/10.3390/ijerph181910106

Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research. *Psychological Methods*, *21*(4), 475–492. https://doi.org/10.1037/met0000081

Meade, B., & Castle, T. (2021). "It's What I Do that Defines Me": Real Life Superheroes, Identity, and Vigilantism. *Deviant Behavior*, 1–15. https://doi.org/10.1080/01639625.2021.1953948

Miller, J. D., Hoffman, B. J., Gaughan, E. T., Gentile, B., Maples, J., & Keith Campbell, W. (2011). Grandiose and vulnerable narcissism: A nomological network analysis. *Journal of Personality*, *79*(5), 1013–1042. https://doi.org/10.1111/j.1467-6494.2010.00711.x

Oakley, B., Knafo, A., Madhavan, G., & Wilson, D. S. (2011). *Pathological Altruism*. Oxford University Press.

Paramboukis, O., Skues, J., & Wise, L. (2016). An Exploratory Study of the Relationships between Narcissism, Self-Esteem and Instagram Use. *Social Networking*, *5*(2), 82–92. https://doi.org/10.4236/sn.2016.52009

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The Development and Psychometric Properties of LIWC2015*. https://doi.org/10.15781/T29G6Z

Preotiuc-Pietro, D., Carpenter, J., Giorgi, S., & Ungar, L. (2016). Studying the Dark Triad of Personality through Twitter Behavior. *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 761–770. https://doi.org/10.1145/2983323.2983822

Pyżalski, J. (2012). From cyberbullying to electronic aggression: Typology of the phenomenon. *Emotional and Behavioural Difficulties*, *17*(3–4), 305–317. https://doi.org/10.1080/13632752.2012.704319

Raskin, R., Novacek, J., & Hogan, R. (1991). Narcissistic self-esteem management. *Journal of Personality and Social Psychology*, *60*(6), 911–918. https://doi.org/10.1037/0022-3514.60.6.911

Rentzsch, K., & Gebauer, J. E. (2019). On the Popularity of Agentic and Communal Narcissists: The Tit-for-Tat Hypothesis. *Personality and Social Psychology Bulletin*, *45*(9), 1365–1377. https://doi.org/10.1177/0146167218824359

Roller, M. R., & Lavrakas, P. J. (2015). *Applied qualitative research design: A total quality framework approach* (pp. xviii, 398). The Guilford Press.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013a). Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLOS ONE*, *8*(9), e73791. https://doi.org/10.1371/journal.pone.0073791

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013b). Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLOS ONE*, *8*(9), e73791. https://doi.org/10.1371/journal.pone.0073791

Shanahan, N., Brennan, C., & House, A. (2019). Self-harm and social media: Thematic analysis of images posted on three social media sites. *BMJ Open*, *9*(2), e027006. https://doi.org/10.1136/bmjopen-2018-027006

Speckmann, F. (2021). Web scraping: A useful tool to broaden and extend psychological research. *Zeitschrift Für Psychologie*, *229*(4), 241. https://doi.org/10.1027/2151-2604/a000470

Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012). Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistic Analysis of Tweets. *2012 11th International Conference on Machine Learning and Applications*, *2*, 386–393. https://doi.org/10.1109/ICMLA.2012.218

Szymczak, P., Sawicki, A., & Jaworski, M. (2020). How narcissists see the social world? Trust, cynicism, and trifurcated model of narcissism. *Current Psychology*. https://doi.org/10.1007/s12144-020-01215-z

Tuchband, A. (2018, April). *Impact of Social Media: A Thematic Analysis exploring body image and self-esteem amongst 18-24 year old women* [Other]. Manchester Metropolitan University. https://e-space.mmu.ac.uk/621701/

Underberg, J. E., Gollwitzer, A., Oettingen, G., & Gollwitzer, P. M. (2020). The Best Words: Linguistic Indicators of Grandiose Narcissism in Politics. *Journal of Language and Social Psychology*, *39*(2), 271–281. https://doi.org/10.1177/0261927X19883898

Utama, A. A. G. S., Soewarno, N., & Agustia, D. (2020). *Accounting Executives Narcissism: A Review and Research Agenda Using Content Analysis* (SSRN Scholarly Paper No. 3687828). Social Science Research Network. https://papers.ssrn.com/abstract=3687828

Yang, Z., Sedikides, C., Gu, R., Luo, Y. L. L., Wang, Y., Yang, Y., Wu, M., & Cai, H. (2018). Communal narcissism: Social decisions and neurophysiological reactions. *Journal of Research in Personality*, *76*, 64–73. https://doi.org/10.1016/j.jrp.2018.07.003

Yarkoni, T. (2010). Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363–373. https://doi.org/10.1016/j.jrp.2010.04.001

Zajenkowski, M., & Szymaniak, K. (2021). Narcissism between facets and domains. The relationships between two types of narcissism and aspects of the Big Five. *Current Psychology*, *40*(5), 2112–2121. https://doi.org/10.1007/s12144-019-0147-1

Zeigler-Hill, V., Sauls, D., Ochoa, V., Kopitz, J., & Besser, A. (2021). Narcissism and Motives to Pursue Status Through the Use of Dominance-Based Strategies, Prestige-Based Strategies, and Leadership-Based Strategies. *Evolutionary Psychological Science*, *7*(3), 254–272. https://doi.org/10.1007/s40806-021-00278-w

6. Thesis discussion

6.1. Summary

The thesis began with the development of a theoretical description of narcissism, aggression and narcissistic aggression, framing narcissistic aggression as fundamentally a defensive strategy to protect fragile narcissistic beliefs that are easily contradicted and threatened by experience (Baumeister et al., 2000; Bushman & Baumeister, 1998). This served as a theoretical framework guiding the design and interpretation of the studies described in the thesis, and provided an explicit map of the theoretical phenomenon of narcissistic aggression as understood by this thesis. Insofar as the theoretical account of narcissistic aggression developed by this comprehensive review represents an outcome of the thesis, it is compatible with the most cohesive contemporary descriptions available in the research literature, which report a strong link between narcissism and aggression mediated by interpersonal provocation (Du et al., 2021; Kjærvik & Bushman, 2021).

The following empirical chapters described a developing sequence of exploratory studies applying different methodologies to investigate narcissistic aggression and relevant subtype differences. These studies can be summarised in terms of several key results. Grandiose narcissism was observed to be associated with more of a 'dispositional' inclination towards aggression that was not necessarily motivated strongly by situational factors. This may indicate grandiose narcissistic aggression is used as an assertive strategy for proactively controlling an individual's social context, that is fundamentally self-motivated and less influenced by the actions of others (O'Reilly & Hall, 2021). Whilst certainly antisocial as a behaviour, this form of grandiose narcissistic aggression may represent a disposition towards less overtly maladaptive aggressive behavioural strategies (Kampe et al., 2021; Loeffler et al., 2020) and better mental health (Loeffler et al., 2020; Rose, 2002). In contrast, vulnerable narcissistic aggression was characterised as more of a classically defensive

threat response featuring hostile reactive behaviours, coherent with the association between vulnerable narcissism and reactive aggression evident in the wider literature (Du et al., 2021; Kjærvik & Bushman, 2021), and maladaptive or avoidant behavioural strategies more generally (Dickinson & Pincus, 2003; Kampe et al., 2021). This difference in controlled and proactive versus reactive strategizing is consistent with research delineating grandiose and vulnerable narcissistic aggression in terms of emotional stability (Czarna et al., 2021; Zajenkowski et al., 2018). More fundamentally, this might represent the aggressive behavioural manifestation of characteristic subtype differences in introversion and extraversion (Jauk et al., 2017).

When examining cognitive processes preceding aggressive behaviour, both grandiose and vulnerable narcissism were associated with increased threat sensitivity independently of whether the content of a situation was threatening or not, coherent with the broader research literature (Du et al., 2021; Hart et al., 2017; Kjærvik & Bushman, 2021, Rasmussen, 2016). However, humiliation and negatively-valenced perceptions of otherwise ambiguous situations were suggested to feature more strongly in vulnerable narcissistic threat detection, highlighting this distinction between subtypes. It may be that ambiguity allows hostile and negative narcissistic preconceptions associated with vulnerable narcissism (Miller et al., 2011) to project greater influence over the interpretation of a potentially threatening situation, without the potentially conflicting evidence of unambiguous features. This creates the conditions for perceiving deliberate acts of rejection and ego-threat necessary for narcissistic aggression, and confirming vulnerable narcissistic beliefs (Hart et al., 2017; Kim & Barry, 2021), which may be amplified and justified by a predisposition to frustrated feelings of humiliation. In contrast, grandiose narcissism may be less concerned with this type of hostile rumination and associated more exclusively with straightforward unambiguous threats (Hart et al., 2017; O'Reilly & Hall, 2021). This highlights the richness of situational factors in narcissistic aggression, and the relevance of situational subtype differences. Whilst these effects were clearly evident in psychometric trait measures, they were less evident in pupil dilation as an exploratory physiological measure of threat perception, or associated behavioural

measures. This may indicate that narcissistic aggression is more than a sum of its parts, and that the essential influence of more nuanced personal and situational factors means that narcissistic behaviour is not reducible past a certain threshold.

When exploring narcissistic behaviour (including narcissistic aggression) in the naturalistic context of an online community, subtype differences were observed to be less pronounced. The expression of vulnerable narcissistic beliefs was facilitated whilst apparently not coinciding with archetypally avoidant behaviours, whereas grandiose narcissistic behaviour was clearly influenced by superficially prosocial considerations, in contradiction with archetypal grandiose agency. This contributes to the growing body of evidence suggesting the agentic narcissism may be less prevalent in communal contexts than the overwhelming influence of agentic narcissism in the literature might lead one to expect (Joyce et al., 2019; Rentzsch & Gebauer, 2019. Narcissism in communal contexts is indeed apparently better described by a distinct communal subtype of narcissism (Gebauer et al., 2012; Gebauer & Sedikides, 2018), featuring both vulnerable and grandiose aspects. More generally, this highlights that narcissistic behaviour in community-moderated contexts is heavily mediated by forces imposed by the community, and may not be reflected by classical trait-based descriptions of narcissistic agency. Importantly, this mediation is more than a simple limitation on narcissistic behaviour, but rather is better described as the facilitation of certain emphasised and socially sanctioned narcissistic behaviours. Aggression in these communities is not well understood, but may be characterised by ‘pathological altruism’ wherein aggressive behaviours are justified by the prosocial but ultimately self-interested motivation to gain community acceptance (Kaufman & Jauk, 2020). Finally, whilst trait differences were observed to be less clearly delineated in a qualitative account of narcissistic behaviour, quantitative trait differences were nevertheless conserved as word frequency-derived metrics when examining the same dataset using text analysis. This possibly represents broad differences in introversion and extraversion evident in the qualitative account (Jauk et al., 2017). Regardless, this suggests that trait differences can be meaningfully retained in qualitative datasets, but that these traits may not be as cleanly separated in practice as they are in theory.

6.2. Contributions

The results described above represent theoretical contributions that advance the conceptual understanding of narcissistic aggression. There are several general theoretical contributions to note. The primary contribution, and that most directly related to the overall topic of the thesis, is furthering the understanding of grandiose and vulnerable subtype differences in narcissistic aggression as they occur in response to the same threatening stimulus. This provides a more direct comparison of these subtypes than that available in numerous otherwise comprehensive review articles (Du et al., 2021; Kjærvik & Bushman, 2021), which may prove useful for those looking for an account of narcissistic aggression where one can be confident the operationalisations of narcissism and aggression featured are consistent.

However, the thesis also provides more practical methodological contributions, that advise the approaches future research should take and provide perspective on the validity of previously used approaches. These methodological contributions are slightly more diverse. The distinguishing methodological feature of the studies described in the thesis is the use of exploratory methods to innovate and improve upon existing approaches, or at least expand the existing range of approaches with a new alternative that might provide different insight. The success of several of these exploratory methods should be taken as encouraging validation of future research in this area taking a similarly exploratory approach, providing a practical proof of concept demonstrating the potential for generating new insight that these methods provide. Perhaps more importantly, these methods also contribute a pertinent reminder that insight can also be generated through failure, as where these methods were unsuccessful in this thesis they nevertheless had something meaningful to say about whatever they intended to measure, that was visible only in the relief cast by this failure. For example, whilst the word frequency-derived metric of grandiose narcissism featured in the fourth empirical chapter was unsuccessful in its intended purpose, its failure illustrated an unexpected and theoretically important separation between psychometric-derived and qualitative-derived accounts of narcissistic traits.

Whilst this type of insight is wholly legitimate and should not be considered a concession, it also necessarily goes hand-in-hand with the more straightforward insight that when an exploratory method is unsuccessful it demonstrates that the more 'tried-and-tested' alternatives to exploratory methods are as widely used as they are for a reason. For example, the accessible, non-controversial and conceptually simple operationalisations of aggression discussed critically earlier in this thesis have legitimate limitations, but it can also be easy to take these methods for granted given the greater level of scrutiny it is possible to subject them to. This is arguably an obvious observation, but as is the case with many obvious assumptions in science, without practical experience to illustrate its truth it becomes easily overlooked. Therefore, both the successes and failures of the exploratory methods featured in this thesis should be interpreted as contributions, encouraging confidence in both the use of exploratory methods and popular methods validated by existing research.

Perhaps the most striking methodological contribution of the thesis is the demonstration of the successful use of qualitative methods to investigate narcissistic behaviour in a naturalistic dataset, including narcissistic aggression. This contribution builds upon the small but growing lineage of research in this area, supporting not only the unique insights gained from studying narcissism in online communities (Balcerowska & Sawicki, 2022; Joyce et al., 2019) but also the importance of doing so (Bogolyubova et al., 2018; Brailovskaia et al., 2020), and provides guidance and encouragement for an approach which otherwise has little presence in the contemporary research literature. The thesis demonstrated both a purely qualitative thematic analysis approach, and a hybrid qualitative-quantitative approach integrating thematic analysis and word frequency data. Importantly, the contribution is not merely arbitrarily demonstrating a means in which qualitative methods can be applied to an area studied traditionally with quantitative methods, but rather demonstrating that qualitative methods offer the potential for their own original insights in this area. For example, an individual may exhibit high trait narcissism on a psychometric scale, but the practical implications of this high trait score in terms of how the individual behaves and communicates may be more particular in ways only visible through the

detailed and highly individualised lens of qualitative analysis (Henwood & Pidgeon, 1992). Naturally, narcissism (like many psychological phenomena) owes its existence and much of its fundamental characterisation to early qualitative research (Levy et al., 2011), but perhaps this historical approach to investigating narcissism has become 'lost in the weeds' of contemporary experimental psychology. A contribution of the thesis is therefore providing evidence to suggest that up-to-date qualitative accounts of narcissistic behaviour should not be overlooked, and may help keep the contemporary conception of narcissism up-to-date and representative of changing social contexts. The increasing prevalence and interest of communal narcissism (Gebauer & Sedikides, 2018) represents a conceptual step in the right direction, as the qualitative results of the thesis suggests community factors may be an influential force mediating the expression of narcissistic traits. More intuitively, as the social environments many inhabit become hypersocial, highly connected and highly influenced by participation in massive global communities where individuals must navigate and compromise around the power of the collective. Narcissism is situated in the interesting position of having to engage with apparently non-narcissistic collective behaviour, whilst more covertly being facilitated by the forms of communication dominating this collective behaviour, which are highly superficial and exploitable by narcissistic behaviour and cognition (Balcerowska & Sawicki, 2022; Buffardi & Campbell, 2008). This highlights that communal narcissism is not merely a product of limits on the manifestation of narcissistic traits but also a product of community factors that may emphasise and facilitate particular aspects of narcissism.

Another key methodological contribution of the thesis is the demonstration of a simple means of accessing a massive naturalistic dataset via webscraping. Webscraping allowed the actual process of data collection involved in the qualitative study to be completed in perhaps an hour, requiring only intermediate computer literacy and knowledge of the highly accessible Python coding language. Importantly, the speed of this approach led to no specific decline in data quality, with the exported data able to be highly structured and filtered if necessary. Of course, the thesis also

demonstrates that a qualitative dataset of this nature does not necessitate an exclusively qualitative approach to analysis. The text analysis method demonstrated provided a useful quantitative metric allowing the exploration of narcissistic traits, which should encourage the use of qualitative datasets of this nature even by researchers who are interested in explicitly quantitative questions or who are otherwise unable to use qualitative methods. Naturally, the ability to use text analysis data to explore narcissistic traits and linguistic behaviours depends on there being access to some measure of trait narcissism in the dataset. Whilst a combination of past research (Holtzman et al., 2019) and qualitatively-derived metrics allowed for a measure of trait narcissism in the qualitative chapter of the thesis, this approach may not be applicable to other research. As a result, a major contribution of the thesis to future research taking a similar approach is the development of qualitatively-derived word frequency metrics of narcissistic traits, that could be used as an operationalisation of these traits in future purely quantitative text analysis research. This would represent a significant expansion of the existing word frequency metrics available, which focus almost exclusively on a 'dark triad' conception of grandiose narcissism (Bogolyubova et al., 2018; Ireland & Mehl, 2014; Yarkoni, 2010). The thesis therefore both demonstrates the success of this method and directly facilitates the application of this method to future research.

More prosaic, but no less significant, other important methodological contributions of the thesis include validating the application of several less frequently-used psychometric measures. The Five Factor Narcissism Inventory (FFNI; Sherman et al., 2015), which was described and judged to be an appropriate and promising metric at the beginning of the empirical component of the thesis, was shown to be a useful measure with good construct validity, insofar as it had strong, theoretically coherent and consistent relationships with other related psychometric and behavioural measures. The MacLean Screening Instrument for Borderline Personality Disorder (Zanarini et al., 2003), introduced in the second empirical chapter, was similarly shown to be a useful measure of more reactive and unstable traits underlying narcissistic aggression, in particular vulnerable narcissistic aggression (Kampe et al.,

2021). Whilst this should naturally be considered an indirect measure and ideally used alongside another operationalisation of aggression, the encouraging consistency and coherency of this measure demonstrated by the thesis should validate the use of this measure in future research. Lastly, using vignettes to operationalise specific aspects of narcissistic threat perception related to hostility and humiliation was shown to be successful. Responses to vignettes were coherent with relevant trait measures, and sensitive to both situational and individual differences, supporting previous research adopting this approach (Hart et al., 2017). The thesis therefore contributes much needed validation of the FFNI, a promising initial example of the application of borderline personality disorder traits as a psychometric operationalisation of narcissistic aggression, and a demonstration of vignettes as an operationalisation of narcissistic threat perception.

Finally, without detracting from the numerous direct contributions of the thesis, it is diligent to point out that a contribution of the thesis is also the much maligned but necessary insight provided by null findings, which help to guide future research away from particularly challenging, impractical or invalid questions and methods. The thesis contributes cautionary reflection on the widespread use of cyberball as an operationalisation of aggression, attributing null findings in part due to the suspected incredulity of participants. The thesis also suggests operationalisations of narcissistic aggression that move away from self-report measures should be approached cautiously, with the use of pupillometry and an associated behaviour task producing interesting but challenging results. Whilst these results justify further exploration, it is not possible to confidently recommend the use of these measures over less exotic and more established psychological methods.

6.3. Limitations

The limitations of the thesis are straightforward and largely summarised as the limited success of some of the exploratory methods used, as described above, and most notably the mixed results of the conflict detection method featured in the third

empirical chapter. The difficulties of these methods may be attributable to several reasons that are discussed in the relevant empirical chapters, but fundamentally this limitation was implicitly acknowledged before beginning work on the empirical component of the thesis, as a necessary consequence of adopting an exploratory approach and venturing off the beaten path. The promise of this approach was generating fresh insight or gaining a different perspective not offered by established methods, which remains a worthwhile and important choice, but the possibility that this approach might nevertheless yield dead ends was recognised. This freedom to explore new questions and embrace the uncertainty of new ideas in spite of the risks is a necessary part of scientific activity, and one that should be encouraged. However, scientific activity should also be concerned with being useful where possible, and so one must always be careful to balance creativity with utility, and avoid taking unnecessary risks.

A more significant limitation that could not be anticipated is the impact of the COVID-19 pandemic on the broader research project represented by the thesis. Whilst relatively fortunate in this regard, as the bulk of data collection was performed prior to the pandemic, it should be acknowledged that it nevertheless resulted in a significant practical limitation. Specifically, the cessation of lab-based research meant a planned neuroimaging study that would have constituted the fourth empirical chapter had to be cancelled, resulting in the loss of time spent preparing for that study, and a deviation from the original planned structure of the project which featured neuroscientific investigation of narcissistic aggression more prominently. Instead, the cancellation of this planned study prompted the pivot to the qualitative study documented in the fourth empirical chapter presented in the thesis, which was arguably justified by the limited success of the approach taken in the third empirical chapter. This carried on the vein of exploratory approaches to investigating narcissistic aggression, and could be conducted without any imposition from pandemic-related limitations. On reflection, this is a ‘happy accident’, as without the qualitative study the thesis would have lost some of its most interesting theoretical

and methodological contributions. Nevertheless, it is important to acknowledge that this study was bought with the loss of another.

References

Balcerowska, J. M., & Sawicki, A. J. (2022). Which aspects of narcissism are related to Social Networking Sites addiction? The role of self-enhancement and self-protection. *Personality and Individual Differences*, *190*, 111530. https://doi.org/10.1016/j.paid.2022.111530

Baumeister, R. F., Bushman, B. J., & Campbell, W. K. (2000). Self-Esteem, Narcissism, and Aggression: Does Violence Result From Low Self-Esteem or From Threatened Egotism? *Current Directions in Psychological Science*, *9*(1), 26–29. https://doi.org/10.1111/1467-8721.00053

Bogolyubova, O., Panicheva, P., Tikhonov, R., Ivanov, V., & Ledovaya, Y. (2018). Dark personalities on Facebook: Harmful online behaviors and language. *Computers in Human Behavior*, *78*, 151–159. https://doi.org/10.1016/j.chb.2017.09.032

Buffardi, L. E., & Campbell, W. K. (2008). Narcissism and Social Networking Web Sites. *Personality and Social Psychology Bulletin*, *34*(10), 1303–1314. https://doi.org/10.1177/0146167208320061

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology*, *75*(1), 219–229. https://doi.org/10.1037/0022-3514.75.1.219

Czarna, A. Z., Zajenkowski, M., Maciantowicz, O., & Szymaniak, K. (2021). The relationship of narcissism with tendency to react with anger and hostility: The roles of neuroticism and emotion regulation ability. *Current Psychology*, *40*(11), 5499–5514. https://doi.org/10.1007/s12144-019-00504-6

Dickinson, K. A., & Pincus, A. L. (2003). Interpersonal analysis of grandiose and vulnerable narcissism. *Journal of Personality Disorders*, *17*(3), 188–207. https://doi.org/10.1521/pedi.17.3.188.22146

Du, T. V., Miller, J. D., & Lynam, D. R. (2021). The relation between narcissism and aggression: A meta-analysis. *Journal of Personality*, *n/a*(n/a). https://doi.org/10.1111/jopy.12684

Gebauer, J. E., & Sedikides, C. (2018). Communal Narcissism: Theoretical and Empirical Support. In A. D. Hermann, A. B. Brunell, & J. D. Foster (Eds.), *Handbook of Trait Narcissism: Key Advances, Research Methods, and Controversies* (pp. 69–77). Springer International Publishing. https://doi.org/10.1007/978-3-319-92171-6_7

Gebauer, J. E., Sedikides, C., Verplanken, B., & Maio, G. R. (2012). Communal narcissism. *Journal of Personality and Social Psychology*, *103*(5), 854–878. https://doi.org/10.1037/a0029629

Hart, W., Adams, J. M., & Tortoriello, G. (2017). Narcissistic responses to provocation: An examination of the rage and threatened-egotism accounts. *Personality and Individual Differences*, *106*, 152–156. https://doi.org/10.1016/j.paid.2016.10.049

Henwood, K. L., & Pidgeon, N. F. (1992). Qualitative research and psychological theorizing. *British Journal of Psychology*, *83*(1), 97–111. https://doi.org/10.1111/j.2044-8295.1992.tb02426.x

Holtzman, N. S., Tackman, A. M., Carey, A. L., Brucks, M. S., Küfner, A. C. P., Deters, F. G., Back, M. D., Donnellan, M. B., Pennebaker, J. W., Sherman, R. A., & Mehl, M. R. (2019). Linguistic Markers of Grandiose Narcissism: A LIWC Analysis of 15 Samples. *Journal of Language and Social Psychology*, *38*(5–6), 773–786. https://doi.org/10.1177/0261927X19871084

Ireland, M. E., & Mehl, M. R. (2014). Natural language use as a marker of personality. In *The Oxford handbook of language and social psychology* (pp. 201–218). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199838639.001.0001

Jauk, E., Weigle, E., Lehmann, K., Benedek, M., & Neubauer, A. C. (2017). The Relationship between Grandiose and Vulnerable (Hypersensitive) Narcissism. *Frontiers in Psychology*, *8*, 1600. https://doi.org/10.3389/fpsyg.2017.01600

Joyce, L., Chrysosferidis, J., Mousavi, S., Barry, C., & Benson, C. (2019). Perceptions of Agentic and Communal Narcissism on Facebook. *Cyberpsychology, Behavior, and Social Networking*. https://doi.org/10.1089/cyber.2019.0135

Kampe, L., Bohn, J., Remmers, C., & Hörz-Sagstetter, S. (2021). It's Not That Great Anymore: The Central Role of Defense Mechanisms in Grandiose and Vulnerable Narcissism. *Frontiers in Psychiatry*, *12*, 661948. https://doi.org/10.3389/fpsyt.2021.661948

Kaufman, S. B., & Jauk, E. (2020). Healthy Selfishness and Pathological Altruism: Measuring Two Paradoxical Forms of Selfishness. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/article/10.3389/fpsyg.2020.01006

Kim, H., & Barry, C. T. (2021). The moderating effect of intolerance of uncertainty on the relation between narcissism and aggression. *Current Psychology*. https://doi.org/10.1007/s12144-021-01542-9

Kjærvik, S. L., & Bushman, B. J. (2021). The link between narcissism and aggression: A meta-analytic review. *Psychological Bulletin*, *147*(5), 477–503. https://doi.org/10.1037/bul0000323

Levy, K. N., Ellison, W. D., & Reynoso, J. S. (2011). A Historical Review of Narcissism and Narcissistic Personality. In *The Handbook of Narcissism and Narcissistic Personality Disorder* (pp. 1–13). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118093108.ch1

Loeffler, L. A. K., Huebben, A. K., Radke, S., Habel, U., & Derntl, B. (2020). The Association Between Vulnerable/Grandiose Narcissism and Emotion Regulation. *Frontiers in Psychology*, *11*, 519330. https://doi.org/10.3389/fpsyg.2020.519330

Miller, J. D., Hoffman, B. J., Gaughan, E. T., Gentile, B., Maples, J., & Keith Campbell, W. (2011). Grandiose and Vulnerable Narcissism: A Nomological Network Analysis. *Journal of Personality*, *79*(5), 1013–1042. https://doi.org/10.1111/j.1467-6494.2010.00711.x

O'Reilly, C. A., & Hall, N. (2021a). Grandiose narcissists and decision making: Impulsive, overconfident, and skeptical of experts–but seldom in doubt.

Personality and Individual Differences, 168, 110280. https://doi.org/10.1016/j.paid.2020.110280

O'Reilly, C. A., & Hall, N. (2021b). Grandiose narcissists and decision making: Impulsive, overconfident, and skeptical of experts–but seldom in doubt. *Personality and Individual Differences, 168,* 110280. https://doi.org/10.1016/j.paid.2020.110280

Rasmussen, K. (2016). Entitled vengeance: A meta-analysis relating narcissism to provoked aggression. *Aggressive Behavior*, *42*(4), 362–379. https://doi.org/10.1002/ab.21632

Rentzsch, K., & Gebauer, J. E. (2019). On the Popularity of Agentic and Communal Narcissists: The Tit-for-Tat Hypothesis. *Personality and Social Psychology Bulletin*, *45*(9), 1365–1377. https://doi.org/10.1177/0146167218824359

Rose, P. (2002). The happy and unhappy faces of narcissism. *Personality and Individual Differences*, *33*(3), 379–391. https://doi.org/10.1016/S0191-8869(01)00162-3

Sherman, E. D., Miller, J. D., Few, L. R., Campbell, W. K., Widiger, T. A., Crego, C., & Lynam, D. R. (2015). Development of a Short Form of the Five-Factor Narcissism Inventory: The FFNI-SF. *Psychological Assessment*, *27*(3), 1110–1116. https://doi.org/10.1037/pas0000100

Yarkoni, T. (2010). Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363–373. https://doi.org/10.1016/j.jrp.2010.04.001

Zajenkowski, M., Maciantowicz, O., Szymaniak, K., & Urban, P. (2018). Vulnerable and Grandiose Narcissism Are Differentially Associated With Ability and Trait Emotional Intelligence. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/article/10.3389/fpsyg.2018.01606

Zanarini, M. C., Vujanovic, A. A., Parachini, E. A., Boulanger, J. L., Frankenburg, F. R., & Hennen, J. (2003). A screening measure for BPD: The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD). *Journal of Personality Disorders*, *17*(6), 568–573. https://doi.org/10.1521/pedi.17.6.568.25355

Reflective Conclusion

I am often struck by the promise and potential of psychology as a field of study. It is concerned with understanding the parts of the natural world that are most intimate to us and connect most directly to the workings of our human world. More than any other science, psychology is the study of everyday and mundane phenomena that govern the course of our lives whilst evading our understanding, and arguably has the most to say about the aspects of our experience that we find most intuitively interesting or concerning. Psychological research should be a productive activity with an important place in our society, helping us to dispel the mysteries of the poorly understood minds we inherit, and transforming knowledge into practical advice on how to interact with- and manage one another in a more prosocial, productive and evidenced way. It was this belief that motivated me to embark on my doctoral studies, which I viewed as an important first step on the path to a career through which I might eventually make contributions to science and improve the way we live.

Whilst my interest and appreciation for psychology at the close of my studies is a lot broader, at the beginning it was very specific, and driven by a specific ambition concerning how this belief in the importance and utility of psychological research could be harnessed. It is hopefully evident that this interest was in studying the risk-factors that cause healthy (i.e. non-clinical) individuals to engage in aggressive behaviour, and what psychological research can tell us about how to predict and manage aggression in our society. This interest in aggression came from a profound personal concern that there was something confused or backwards in our cultural attitude towards aggressive behaviour, and in particular the way we regard individuals who suffer from serious but non-clinical issues with controlling their aggression. I was struck by how much literature there was on the topic, and yet how challenging it was to be confident about why, when and how individuals become aggressive. I have since come to appreciate why that is the case, and why my initial assessment was in many ways an uneducated assessment of the substantial work of those whose insights into aggression I had taken for granted. But there nevertheless

remains some truth to this observation, and I felt a strong motivation to get involved and see if I couldn't generate some more definite answers through my own research. Otherwise, it felt inevitable that without any clear direction from science, our society's best solution to issues of aggression would continue to be (both literally and figuratively) putting it in a box and forgetting about it.

This thesis is a record of my attempts to understand some small piece of the intimidating question of why, when and how individuals become aggressive. My approach in my doctoral studies was to return to basics and build my own theory first before beginning to imagine any way of applying it to generate social impact. Having completed my doctoral studies, what I have achieved are a series of exploratory studies that build some firm, evidence-based theoretical foundations for asking further questions about aggression and its relationship with the risk factor in which I was particularly interested - narcissism - which hopefully has been demonstrated to be an interesting and dynamic personality trait well-worthy of special interest in the context of aggression. I am fortunate and encouraged that my work was largely successful, and where it was not successful produced meaningful insight regardless. As with all research, the studies documented in this thesis represent something of a 'tip of an iceberg' and belie hundreds of hours of design and data collection, and three years of reflection and personal development, which in some cases yielded professional and academic insights equally as impactful as the data they eventually generated. I personally acknowledge these alongside the more formal outcomes of my doctoral studies as markers of my education and development as a researcher. Perhaps the best way to summarise the overall character of these insights is a sort of unromantic appreciation for the philosophy of psychology, where it can be difficult to disentangle the phenomena being studied from the methods used to study them, and where a researcher may have to reconcile their work with the observation that to an extent the phenomena they study are abstractions that do not have a wholly objective existence. Personality traits and lab-based operationalisations of aggression are particularly interesting and challenging examples of these kinds of phenomena, being always obfuscated by necessary but influential methodological decisions on the part

of the researcher, and always several degrees separated from the intended object of study. Far from being disorienting or demotivating, these insights are empowering in their ability to demystify these phenomena by encouraging embracing their inherently mysterious nature. Similarly, far from leading to blissfully unconcerned navel-gazing, tackling the implications of these insights head-on leads to a pragmatic, diligent and reflective approach that can only be beneficial to research practice. Although I have a long way to go towards personally applying this approach to the extent described here, I feel I have earned this initial but immensely valuable lesson into what it means to be a researcher and to practise the science of psychology.