Irena Pilch[*]
Magdalena Hyla[*]

Narcissism and self-esteem revisited: The relationships between the subscales of the NPI and explicit/implicit self-esteem

Abstract: *The results of the previous research which tried to investigate links between self-esteem and narcissism in search of the source of narcissistic fragility are not consistent. The aim of the study was to contribute to the understanding of this complex relation by assessing relationships between the four facets of grandiose narcissism measured by the NPI and the two kinds of self-esteem i.e. explicit (ESE) and implicit (ISE), and by comparing the results with theoretical models. The Implicit Relational Assessment Procedure, a more recent measure based on response latency, was used to assess ISE. The analysis confirms the reports on NPI heterogeneity. Different relationships between the particular facets of narcissism and ESE/ISE were observed. These differences indicate that composite measures of narcissism and ISE do not reflect the entire complexity of the phenomena. The results provide support for the two theoretical models i.e. the global marker model and the mask model.*

Key words: *self-esteem, narcissism, Narcissistic Personality Inventory, implicit measures, implicit self-esteem, IRAP*

Introduction

Narcissism as a trait

From the perspective of social and personality psychology narcissism is a dimension characterising "normal" variability within the personality. Narcissistic individuals are characterised by egocentrism/self-absorption, a sense of superiority and entitlement, motivation for self-enhancement and self-promotion regardless of the costs (Farwell & Wohlwend-Lloyd, 1998; Sedikides & Gregg, 2001; Morf & Rhodewalt, 2001). This type of narcissism, defined also as grandiose narcissism, is related to extraversion, self-confidence, emotional resilience, dominance, and agency, but also to a lack of empathy, exhibitionism, manipulativeness, aggressiveness and exploitativeness (see Miller & Campbell, 2008). Grandiose narcissism correlates with positive features of functioning (such as well-being, lack of depression or high self-esteem; Sedikides, Rudich, Gregg, Kumashiro, & Rusbult, 2004, Campbell, Rudich, & Sedikides, 2002; Rose, 2002). Narcissism understood in this way is thus viewed as relatively "adaptive" and differs significantly from clinical narcissism perceived as a disorder.

Vulnerable narcissism is another type of narcissism that is less studied. It is related to introversion, lack of self-confidence, shyness, disposition to experience negative feelings (anxiety, shame) and emotional vulnerability accompanied by egocentrism, grandiose expectations, entitlement and disregard for others (Wink, 1991; Dickinson & Pincus, 2003). Vulnerable narcissists experience greater anxiety and distress in social situations (Besser & Priel, 2009) and are less happy (Rose, 2002). Vulnerable narcissism is directly related to functional impairment and is linked to the grandiose form of narcissism through self-absorption and antagonistic interpersonal style (Miller & Campbell, 2008). The most recent comparative studies confirm that grandiose narcissism and vulnerable narcissism constitute two separate constructs (Given-Wilson, McIllwain, & Warburton, 2011; Luchner, Houston, Walker, & Houston, 2011; Lannin, Guyll, Krizan, Madon, & Cornish, 2014; Miller, Gentile, Wilson, & Campbell, 2013).

The vast majority of studies on narcissism as a trait are conducted using the Narcissistic Personality Inventory (NPI; Raskin & Hall, 1979; 1981) which measures mainly grandiose aspects of narcissism. Thus important features

---

[*]   **University of Silesia, Grażyńskiego 53, 40-126 Katowice, Poland**

Corresponding author: Irena Pilch

of vulnerable narcissism are excluded from the analysis (Cain, Pincus, & Ansell, 2008), although some of them are described with the use of this measure. Despite enormous popularity of the tool, its application is not problem-free. An unstable factor structure of NPI (from two to seven factors, Ackerman, Witt, Donnellan, Trzesniewski, Robins, & Kashy, 2011; in the Polish version – four factors, see Bazińska & Drat-Ruszczak, 2000) and low subscale reliability reported by some researchers (del Rosario & White, 2005; Horton, Bleau, & Drwecki, 2006) result in the analysis which is frequently limited to the total score. However, reducing a complex, multidimensional construct to a single dimension drastically limits the possibilities of result interpretation. Some researchers emphasise that the NPI composite narcissism index includes factors which are more or less adaptive and these factors should be analysed separately. It particularly applies to the Exploitativeness/Entitlement factor (see Emmons, 1984), which correlates negatively with self-esteem and positively with depression or anxiety and is thereby associated with psychological dysfunction (Ruiz, Smith, & Rhodewald, 2001). The most recent NPI validation (Ackerman et al., 2011) has shown that only Leadership exhibits a pattern of correlation which enables Leadership to be classified as adaptive or normal narcissism. The other two facets (i.e. Grandiose/Exhibitionism and Entitlement/Exploitativeness) were categorised as maladaptive with particular emphasis placed on "socially toxic" aspects of the NPI expressed by Entitlement.

Explicit and implicit self-esteem

Self-esteem (SE), one of the most important, but at the same time one of the most controversial psychological constructs, is predominantly understood as global evaluation of the self (Bosson & Swann, 2009). Self-reported SE is frequently studied in both clinical (depression) and non-clinical (well-being) contexts. High explicit self-esteem (ESE) is associated with a definitively positive influence on a number of aspects of an individual's life (Tennen & Affleck, 1993). However, meta-analysis of the results of studies on SE demonstrated that direct benefits related to high SE are limited. Individuals with high ESE are more active, persistent and experience more positive feelings as compared to individuals with low ESE (Baumeister, Campbell, Krueger, & Vohs, 2003). At the same time, a number of relationships between high SE and unwanted events, such as aggression, can be observed (e.g. Baumeister, Smart, & Boden, 1996). In order to explain this phenomenon, it has been postulated that there exist secure high SE (when self-views are realistic and resistant to threat) and fragile high SE (when self-views are more uncertain, vulnerable to challenge and need constant validation) (Kernis, 2003). One of the ways of distinguishing secure from fragile high SE is an analysis of discrepancies between ESE and implicit self-esteem (ISE) (Bosson, Brown, Zeigler-Hill, & Swann, 2003; Jordan, Spencer, Zanna, Hoshino-Browne, & Correll, 2003).

Self-report questionnaires cannot capture implicit self-views or self-views which a particular individual is reluctant to reveal (Buhrmester, Blanton, & Swann, 2011). Introduction of the construct of ISE and indirect measures allowed to capture those elements of self-image which cannot be assessed by traditional measures. ISE can be defined as a construct distinct from ESE, being a result of the action of a separate system of information processing and affecting the behaviour irrespective of reported self-evaluations (Rudman, Dohn, & Fairchild, 2007). This position stems from the dual-process theories of cognition, which assume the existence of two distinct cognitive systems. The first system is fast, automatic, spontaneous, impulsive, associative, unconscious, or implicit, the second one is slow, sequential, systematic, deliberate, reflective, controlled, conscious, or explicit (Nosek, Hawkins, & Frazier, 2011). ISE understood in this way (treated as a result of implicit processes) is defined as an automatic evaluation of the self and is perceived to be inaccessible to conscious processes, whereas ESE is considered as a result of explicit processes.

Another possibility is to conceptualize ISE and ESE as forms of one construct which differ from each other in the techniques of their measurement (indirect or direct; see: Fazio & Olson, 2003). In other words, ISE and ESE can be recognised as separate manifestations of the same core attitude. They can be treated as equal or it can be assumed that ISE, as insusceptible to conscious shaping, is primary (Dijksterhuis, Albers, & Bongers, 2008; see also Olson & Fazio, 2008).

In both cases a strong relationship between ISE and ESE is not expected – both constructs do not correlate with each other or the correlation is poor (Hoffmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Krizan & Suls, 2008; Bosson, Swann, & Pennebaker, 2000; Wierzchon, 2009). The available literature also provides reports indicating that the relationship between ISE and ESE is stronger in individuals who trust their intuitive feelings (Jordan, Whitfield, & Zeigler-Hill, 2007) and mainly in women as compared to men (due to differences in socialisation; Pelham, Koole, Hardin, Hetts, Seah, & DeHart, 2005). This may suggest that people can become aware of their ISE and include it in the creation of self-image.

There is evidence that low ESE or ISE can have similar outcomes. For example, in Greenwald and Farnham (2000) experiment both high ESE and high ISE acted as a buffer protecting an individual in the face of negative experiences. On the other hand, results of other studies demonstrate differences between ESE and ISE perceived as predictors – ISE is a better predictor of spontaneous or emotional behaviour (Bosson et al., 2000; Conner & Barret, 2005; Rudolph, Schroder-Abe, Riketta, & Schutz, 2010).

Possible combinations of ISE and ESE which can be observed in particular individuals have attracted attention of researchers. Despite the fact that both types of SE are formed on the basis of the same events, only ESE is influenced by conscious interpretations and corrections, while only ISE reflects life experiences in a direct manner (Zeigler-Hill, 2006). Due to these differences implicit and explicit self-attitudes can be consistent or inconsistent. Two forms of discrepant SE are possible: discrepant

low SE (damaged SE: low ESE/high ISE; Schroder-Abe, Rudolph, & Schultz, 2007) and discrepant high SE (fragile or defensive SE: high ESE/low ISE; Bosson et al., 2003; Jordan et al., 2003). Both forms of discrepant SE, reflecting lack of integration of self-representations, are considered dysfunctional. They are related to emotional distress, psychological discomfort and lower psychological well-being (Schroder-Abe, Rudolph, & Schultz, 2007; Vater, Schroder-Abe, Schultz, Lammers, & Roepke, 2010). Fragile SE requires constant confirmation or even self-deception. This mechanism is consistent with a conceptualisation of narcissism as a combination of overt grandiosity and unconscious, negative attitudes towards the self (Hovarth & Morf, 2009). On the other hand, damaged SE is associated with numerous unwanted events, such as anger suppression, nervousness, health problems, defensive behaviour and social anxiety disorder (Schroder-Abe, Rudolph, & Schultz, 2007; Schroder-Abe, Rudolph, Wiesner, & Schultz, 2007; Schreiber, Bohn, Aderka, Stangier, & Steil, 2012). Another approach to discrepant low SE is based on the "glimmer of hope" hypothesis, according to which high ISE can have a positive effect on individuals with low ESE, as it can serve as a buffer from unwanted events, thereby counterbalancing negative self-views (Spencer, Jordan, Logel, & Zanna, 2005; Jordan, Logel, Spencer, Zanna, Wood, & Holmes, 2013).

Measurement of ISE and ESE

ISE measures usually do not correlate with one another. This may be the result of complexity and multidimensionality of the construct and the limitations related to measures which are usually characterised by low test-retest reliability (Fazio & Olson, 2003). Currently, the two most commonly used measures of ISE are the Implicit Association Test (IAT; Greenwald & Farnham, 2000) and the Name-Letter Test (NLT; Koole, Dijksterhuis, & van Knippenberg, 2001; Bosson et al., 2003). The IAT is a reaction time task which measures automatic associations of self-relevant and non-self-relevant words with pleasant and unpleasant words (Greenwald & Farnham, 2000). The results reflect the speed of association of positive and negative words with the self. The NLT is based on an assumption that the level of preference for the letters in one's own name reflects the overall self-worth. Buhrmester et al. (2011), in the meta-analysis of the results of studies using the IAT and the NLT, came to the conclusion that neither of them was a valid measure of ISE. According to their opinion, the IAT measures implicit affect processes, while the NLT measures implicit egotism. Their main objections concerned the reduction of SE to simple associations related to the self and being poor predictors of important phenomena, such as well-being and depression. The scores of the IAT and the NLT have modest temporal stability and can be easily influenced using, e.g. subliminal conditioning (Grumm, Nestler, & Von Collani, 2009). Thus the scores can be regarded as a reflection of a state rather than a trait. Although a multiple assessment of ISE would be a solution to determine rather a trait than a state, only Zeigler-Hill (2006) has adopted this approach. One of the disadvantages of the IAT is the necessity to evaluate self-views with reference to other people. However, such an inconvenience is not present in the case of the Single Category IAT (Karpinsky & Steinman, 2006). In conclusion, although the usefulness of SE assessment at the unconscious level is usually not questioned, searching for new and better measures for this construct is postulated.

ISE is commonly measured as a unitary construct, although such an approach is criticised (see Bosson et al., 2000). An attempt to distinguish facets of ISE was undertaken only in a few studies by adding adequate adjectives to the IAT. Campbell et al. studied agency-related and communion-related ISE (Campbell, Bosson, Goheen, Lakey, & Kernis, 2007). Klavina, Schroder-Abe and Schultz (2012) distinguished 4 facets of ISE: self-regard, social, performance and physical SE. Different aspects of SE are tested in a different manner by the Go/No-go Association Test (GNAT; Nosek & Banaji, 2001), which includes division into 4 blocks: self-positive, self-negative, non-self-positive and non-self-negative.

Global ESE is usually measured via self-report questionnaires and treated as a single dimension, however, two component and multiple component approaches have also been proposed (Bosson & Swann; Hyland, Boduszek, Dhingra, & Shevlin, 2014). The most popular measure of global self-esteem is Rosenberg's Self-esteem Scale (Rosenberg, 1965; see also Boduszek, Hyland, Dhingra, & Mallett, 2013).

Narcissism, ESE and ISE

A number of studies have demonstrated positive correlations between grandiose narcissism (as measured by the NPI) and ESE in non-clinical (Bosson, Lakey, Campbell, Zeigler-Hill, Jordan, & Kernis, 2008; Sedikides et al., 2004; Campbell et al., 2007) and clinical samples (Pincus, Ansell, Pimentel, Cain, Wright, & Levy, 2009). Cain et al. (2008, p. 644) confirm that "narcissistic traits assessed by the NPI include adaptive characteristics that are inherent in positive SE" – dominance and self-enhancement. At the same time, as previously mentioned, one of NPI subscales (Exploitativeness/Entitlement) does not correlate with ESE or correlates negatively (which is not taken into account in the studies considering only the total score) and may be regarded as a maladaptive component of narcissism (e.g. Watson, Varnell, & Morris, 1999-2000). In order to separate a positive from a negative aspect of narcissism measured by the NPI, Horton et al. (2006) used self-reported SE as a criterion, regarding the type of narcissism which is not accompanied by high SE as "unhealthy". On the other hand, a vulnerable aspect of narcissism measured by the Hypersensitive Narcissism Scale (Hendin & Cheek, 1997) or the Pathological Narcissism Inventory (Pincus et al., 2009) correlates negatively with ESE in persons with psychiatric disorders and in groups of healthy individuals (Pincus et al., 2009; Miller, Dir, Gentile, Wilson, Pryor, & Campbell, 2010; Maxwell, Donnellan, Hopwood, & Ackerman, 2011; Boldero, Higgins, & Hulbert, 2015).

The belief that explicitly demonstrated high SE is a "false mask" which hides emotionally fragile

self originates from the psychodynamic approach (Kohut, 1976; Kernberg, 1970). One of the postulated sources or manifestations of this phenomenon might be the discrepancy between ESE and ISE exhibited by a narcissistic individual (Bosson et al., 2008; Gregg & Sedikides, 2010). The study results that are consistent with the above mask hypothesis which assumes an interaction between two types of SE were obtained by Jordan et al. (2003) and Zeigler-Hill (2006). The highest level of narcissism was demonstrated by individuals with high ESE and low ISE. Individuals with high levels of both ESE and ISE showed no difference in narcissism from those with low ESE. However, subsequent studies (Campbell et al., 2007; Bosson & Prewitt-Freilino, 2007; Gregg & Sedikides, 2010) did not yield similar results. The mask hypothesis is indirectly supported also by studies which demonstrated a relationship between fragile SE and defensiveness (Jordan et al., 2003) and tendencies towards self-enhancement (Bosson et al., 2003).

Gregg and Sedikides (2010) reported three models explaining mutual relationships between narcissism and ESE/ISE. According to the global marker model, ISE determines ego fragility – narcissism is associated with high ESE and low ISE, and interactions between them are not expected (Gregg & Sedikides, 2010, p. 144). The partial discrepancy model considers one source of ego fragility i.e. high ESE accompanied by low ISE (Gregg & Sedikides, 2010, p. 146). Thus this model corresponds to the mask model described above. Statistically, the partial discrepancy model indicates that ESE will interact significantly with ISE. The third possibility is the full discrepancy model, according to which ego fragility exists not only in the case of high ESE and low ISE (as demonstrated by the partial discrepancy model) but also when low ESE is accompanied by high ISE (damaged SE; Gregg & Sedikides, 2010, p. 145; see also Kernis et al., 2005). This model is different from the global market model, because the former implies that ESE and ISE should interact but should not exert separate main effects on narcissism.

Gregg and Sedikides (2010) obtained the results consistent with the global marker model, however only in the case of ISE measured by the GNAT (not by the IAT). Support for the full discrepancy model was obtained on the basis of the recent studies conducted on a group of individuals diagnosed with narcissistic personality disorder (Vater, Ritter, Schroder-Abe, Schultz, Lammers, Bosson, & Roepke, 2013). The highest level of narcissism was observed in individuals with relatively high ISE and low ESE. The above-mentioned pattern of relationships between ESE and ISE was frequently observed in the groups of depressed patients (e.g. Franck, De Raedt, & De Houver, 2007) and can be characteristic of individuals with the vulnerable form of narcissism. The arguments in favour of the partial discrepancy (mask) model were presented above in this paragraph – this model is the most popular among researchers.

In summary, the study results on the relationships between ESE, ISE and narcissism provide inconclusive or contradictory findings. Bosson et al. (2008) in their meta-analysis did not observe a simple relationship between the NPI and ISE or an interaction between ESE and ISE (as measured by the IAT) as predictors of narcissism. However, they found the relationships for ISE measured by the NLT, which do not support the mask hypothesis. One of the explanations for such a result referred to multidimensionality of narcissism and a possibility that its different dimensions will be differently related to SE. Despite these problems, it can be assumed that the results of empirical research confirm the significance of SE in the study on narcissistic personality.

The current study

There is evidence that high SE of narcissists masks their fragile self and discrepancies between ISE and ESE are one of the possible causes or manifestations of this phenomenon. However, this evidence is inconsistent. Researchers suggest that future research should be done using new measures for determination of ISE and that narcissism should be treated as a multidimensional construct (Bosson et al., 2008; Zeigler-Hill & Besser, 2013). Considering the above recommendations we performed the study, the aim of which was the analysis of relationships between narcissism, its facets (measured by the NPI) and SE. In the study ISE measurement was made using the IRAP, which is a relatively new method of assessing implicit attitudes (Barnes-Holmes, Barnes-Holmes, Power, Hayden, Milne, & Stewart, 2006; Vahey, Nicholson, & Barnes-Holmes, 2015).

Based on the three models proposed by Gregg and Sedikides (2010) which describe potential relations between narcissism and ESE/ISE, the sources of narcissistic fragile self may be found exclusively in low ISE (global marker model), or in the configurations i.e. high ESE-low ISE and high ISE-low ESE (full discrepancy model), or exclusively in the combination of high ESE and low ISE (partial discrepancy model). Each of these three models has some direct and indirect empirical support, described above. Therefore we did not formulate any hypotheses concerning expected relationships between narcissism (total and facet scores) and ISE or concerning the existence and the type of interactions between ESE and ISE in predicting narcissism and its facets. However, positive relationships between ESE and global narcissism with its components were expected, except for the Demand for Admiration subscale (which in the Polish version of the NPI corresponds to some degree to the Exploitativeness/Entitlement; Bazińska & Drat-Ruszczak, 2000).

The analysis of relationships between ESE/ISE and narcissistic personality may contribute to a better understanding of fragile SE construct, which in the past was analysed in the context of narcissism, as this trait may be treated an "ideal indicator of defensiveness" (Jordan et al., 2003, p. 970). Better insight into narcissism as a multidimensional construct, the facets of which may differentiate not only ESE but also ISE and their configurations, may contribute to resolving the doubts related to inconsistent results of the previous studies.

Method

Participants

The current study examined a self-selected sample of university students. Fifty-six persons (7 males) participated as volunteers. They were recruited through an online advertisement that was placed on the websites of the University of Silesia, and the University of Economics, Katowice, Poland. During data analysis the results of 7 participants were excluded from the study. These participants did not get to the IRAP test stage or (despite passing through the practice stage) they obtained the results of less than 80% correct responses in more than one block or the response time was more than 2000 ms in the test stage (see Barnes-Holmes, Barnes-Holmes, Stewart, Boles, 2010, p. 532). Finally, the group consists of forty-nine participants (age $M = 22.0$, $SD = 1.9$).

Procedure

The test was conducted either in a laboratory or in a room developed specifically for the test at the university. The participants had been informed about the procedure and the estimated duration of the task. To combine the results of the questionnaires and the IRAP, the participants were randomly assigned an identification number in order to preserve their anonymity and confidentiality of the data, which they were also informed about. In the first place, the participants completed the section devoted to respondents' particulars, the Rosenberg Self-Esteem Scale (RSES) and the NPI. The next step was based on doing the IRAP task.

Materials

Explicit self-esteem (ESE)

The Polish version of the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965; Łaguna, Lachowicz--Tabaczek, & Dzwonkowska, 2007) was used. It is an unidimensional self-report measure that captures a person's global evaluation of her or his worthiness (for a discussion about the scale's dimensionality see McKay, Boduszek, & Harvey, 2014). The instrument consists of 10 items with the answers on a 4-point scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*) with higher values reflecting higher levels of SE ($\alpha = .86$).

Narcissism

The Polish adaptation of the NPI was applied (Raskin & Hall, 1979, 1981; Bazińska & Drat-Ruszczak, 2000). It is a measure of narcissism most frequently used in subclinical groups. The Polish version of the NPI consists of 34 items with the answers on a 4-point scale ranging from 1 "*it's not me*" to 4 "*it's me*". The test has a four-factor structure i.e. Self-sufficiency ($\alpha = .63$), Vanity ($\alpha = .68$), Leadership ($\alpha = .83$) and Demand for Admiration ($\alpha = .7$).

Implicit self-esteem – IRAP

The IRAP (Barnes-Holmes et al., 2006) is a computer method based on examining response time. It is designed to target stimulus relations by capturing how objects are related to each other according to the participant. Those relations can be understood as implicit attitudes, which are the result of both prior learning history and current contextual variables (Hughes, Barnes-Holmes, & De Houwer, 2011). The IRAP requires participants to respond quickly and accurately to the stimuli which are either consistent or inconsistent with their prior learning history. 50% of the IRAP trials require that participants respond consistently with their learning history and the other 50% of the IRAP trials require inconsistent responses. Due to the fact that the IRAP is based on an assumption that the strength of relationships between stimuli is reflected in the participant's response time, responding should be faster on consistent relative to inconsistent trials. The response time differential between consistent and inconsistent trials provides non-relative index of the strength of implicit attitude towards the measured relationships.

The application of the IRAP as a method of implicit attitude measurement allows to assess the aspects of implicit attitudes that are not assessed by other implicit measures. As opposed to other methods, the result of which is a single metric of overall implicit bias, the IRAP assesses four metrics, one for each of the individual responses that are targeted by the measure (in the current study: D-IRAP-tt1 for the trial type "*I am positive*", D-IRAP-tt2 for the trial type "*I am negative*", D-IRAP-tt3 for the trial type "*I am not positive*", and D-IRAP-tt4 for the trial type "*I am not negative*"). The IRAP can capture these four different stimulus relations independently from one another, which enables a better understanding a history of learning. Researchers are encouraged to interpret these scores (and also their combinations, like D-IRAP-POS for positive traits and D-IRAP-NEG for negative traits in case of the self-esteem IRAP; see Hussey, Thompson, McEnteggart, Barnes-Holmes, & Barnes-Holmes, 2015). Currently, the IRAP is widely used to assess implicit attitudes, such as those related to the self (e.g. Kosnes, Whelan, O'Donovan, & McHugh, 2013).

The validity and reliability of the IRAP as a measure of implicit cognition has been established in previous research (for a review see Golijani-Moghaddam, Hart, & Dawson 2013). Several studies showed that IRAP effects were different for groups of people who would be expected to score differently (e.g. meat-eaters and vegetarians; Barnes-Holmes, Murtagh, Barnes-Holmes, Stewart, 2010). Evidence from IRAP studies indicates that the IRAP has acceptable experimental (e.g. Hussey & Barnes-Holmes, 2012), discriminant (e.g. Power, Barnes-Holmes, Barnes--Holmes, & Stewart, 2009), and predictive validity (e.g. Nicholson & Barnes-Holmes, 2012). Some researchers reported insufficient internal consistency of the IRAP, but the studies which reduced permitted response latency to 2000 ms had acceptable reliability (> .70; Golijani--Moghaddam et al., 2013). A recent meta-analysis of fifteen studies conducted in clinical settings indicated good criterion validity of the IRAP (meta-effect r = .45; Vahey et al., 2015).

In the present study, the original IRAP task was modified to assess ISE. Firstly, the initial set of stimuli (adjectives), reflecting the model of ESE which lies at

the foundation of the RSES, was prepared by a group of university psychologists. After that, these words were rated for pleasantness by a group of university students. The final set of stimuli used to measure ISE (positive targets: good/dobry, capable/zdolny, needed/potrzebny, successful/skuteczny, worthy/wartościowy, satisfied/zadowolony; negative targets: bad/zły, stupid/głupi, useless/bezużyteczny, failure/daremny, worthless/bezwartościowy, sad/smutny) consists of words that were rated as highly pleasant or highly unpleasant.

The stimuli were presented on the monitor screen and the participant's responses were recorded by the IRAP software. The experimenter provided each participant with the instructions included in the experimenter's script (Experimenter script 2.0 from www.irapresearch.org, the Polish-language version available from the authors of the article). The IRAP consisted of blocks in which participants were required to give either a consistent response ("*I am positive*") or an inconsistent one ("*I am negative*"). Based on the findings for the universality of self-positive implicit bias, it was assumed that "I am positive" is a rule consistent with prior learning history (Yamaguchi, Greenwald, Banaji, Murakami, Chen, Shiomura, Kobayashi, Cai, & Krendl, 2007). In the current study, each IRAP involved a minimum of two and a maximum of six practice blocks followed by a fixed set of test blocks. Each block consisted of 24 trials. Each trial consisted of a category label ("*I am*" or "*I am not*") occurring at the top of the screen, one of the 12 target words occurring in the middle of the screen and the two response options, "press 'D' for TRUE" and "press 'K' for FALSE" at the bottom of the screen. All stimuli (label, target and response options) appeared simultaneously (see Appendix A) and displayed on the screen until the participant pressed the key corresponding to the correct response. If an incorrect response was provided, a red 'X' symbol was displayed which was removed after the participant provided the correct response. After providing the correct response, the programme proceeded to the next trial.

Which response was considered correct or incorrect depended on whether "consistent" or "inconsistent" block was administered. The correct response during the "consistent" blocks required the participant to select "TRUE" when "I am" appeared with a positive target stimulus (e.g. good) or "FALSE" when it appeared with negative (e.g. bad). In the same block, when next to "I am not" appeared a positive target stimulus (e.g. good), the correct response was "FALSE" and when next to "I am not" appeared a negative target stimulus, the correct response was "TRUE". The opposite pattern of responding was required during inconsistent blocks. The general rule for responding was alternated across each IRAP block to form three successive pairs of test blocks.

In order to familiarise the participant with the procedure, the IRAP began with a pair of practice blocks which did not differ from the test blocks. The participants were allowed to progress through the test blocks after fulfilling the accuracy (at least 80%) and latency criteria (median latency of less than 2000 ms)

in two successive practice blocks. Failure to meet the criteria resulted in presenting another pair of practice blocks until the participant either achieved the mastery criteria or a maximum of four pairs of practice blocks were completed. If the participant failed to meet the criteria in four pairs of practice blocks, his or her involvement in the experiment was finished and the participant was debriefed, thanked and dismissed. The IRAP test phase began automatically after the completion of a pair of practice blocks and fulfilling the criteria. During the test phase the participants were exposed to six test blocks, alternating between "consistent" and "inconsistent" blocks, each with 24 trials. After each test block, instructions were presented on the screen. They specified the rule applicable in the next block.

Results

Introductory analyses

The IRAP used in this study was similar to the one applied by Drake et al. (Drake, Kellum, Wilson, Luoma, Weinstein, & Adams, 2011). Four D-IRAP scores were calculated, one score for each trial type: D-IRAP-tt1 – Self-Positive: I'm [positive word], D-IRAP-tt2 – Self-Negative: I am [negative word], D-IRAP-tt3 – Self-not Positive: I am not [positive word], D-IRAP-tt4 – Self-not Negative: I am not [negative word]. In our study the results of the analysis of split-half reliability with Spearman-Brown correction demonstrated satisfactory IRAP reliability for individual trial types: Self-Positive $r = .629$, Self-Negative $r = .812$, Self-not Positive $r = .877$, and Self-not Negative $r = .840$.

Firstly, according to the recommendations by Hussey et al. (2015) the D-IRAP-tt2 and D-IRAP-tt3 scores were inverted (multiplied by -1) so that scores > 0 share a common interpretation (i.e. represent positive or not-negative effects). Similarly to previous studies using the IRAP as a measure of SE, D-IRAP-POS score was calculated for positive words (by averaging the Self-Positive and Self-not Positive scores) and D-IRAP-NEG score was calculated for negative words (by averaging the Self-Negative and Self-not Negative scores). The overall D-IRAP score and trial type (D-IRAP-tt) scores were in a positive direction: one-sample t-tests determined that Self-Positive ($t = 11.08$, $p < .001$, Cohen's d = 1.58), Self-not Positive ($t = 3.96$, $p < .001$, Cohen's d = 0.57), Self-not Negative ($t = 4.63$, $p < .001$, Cohen's d = 0.66) and D-IRAP ($t = 7.01$, $p < .001$, Cohen's d =) scores were significantly different from zero. Positive D-scores indicate a self-positive implicit bias. Only Self-Negative ($t = .10$, $p = .9$) scores did not differ from zero, which suggests no bias toward positive or negative self-views.

Table 1 presents the means, standard deviations, and zero-order correlations for the measures in the current study. Mean scores of RSES and NPI subscales did not differ from the scores obtained in other studies using the Polish version of these instruments. The RSES scores correlated only with two NPI subscales i.e. moderately positive with Self-sufficiency and negatively with

Table 1. Means, standard deviations (SD) and zero-order correlations

	1	2	3	4	5	6	7	8	9
1 NPI total	–								
2 Leadership	.81***	–							
3 Demand for Admiration	.70***	.37**	–						
4 Vanity	.56***	.25	.23	–					
5 Self-sufficiency	.54***	.37**	.06	.28*	–				
6 RSES	.13	.17	-.30*	.21	.49***	–			
7 D-IRAP	-.13	.10	-.35*	.09	-.18	.11	–		
8 D-RAP-POS	-.14	.05	-.22	.06	-.29*	-.11	.79***	–	
9 D-IRAP-NEG	-.08	.10	-.34*	.09	-.01	.27	.83***	.30*	–
M	103.1	32.7	30.1	14.9	25.5	30.6	.24	.36	.12
SD	13.9	6.8	6.1	3.7	3.5	4.7	.24	.28	.31

* $p < .05$, ** $p < .01$, *** $p < .001$

Demand for Admiration. The RSES scores and D-IRAP scores were not correlated. Similarly, there were no significant correlations between the RSES and trial-type D-IRAP scores: Self-Positive ($r = -.19$, $p = .18$), Self-not Positive ($r = -.01$, $p = .97$), and Self-not Negative ($r = .09$, $p = .53$), except for a weak positive correlation between the RSES and Self-Negative ($r = .3$, $p = .03$). D-IRAP and D-IRAP-NEG correlated negatively only with Demand for Admiration. Moreover, Self-sufficiency correlated negatively with D-IRAP-POS.

Explicit and implicit self-esteem and their interaction as predictors of narcissism and its facets – regression analyses

In order to determine whether the RSES, the IRAP and their interaction will predict narcissism and its facets, a series of multiple regression analyses was conducted. In every case RSES scores were treated as a predictor, one of the IRAP scores (D-IRAP, D-IRAP-POS or D-IRAP-NEG) was treated as a moderator and one of the NPI scores was treated as a criterion variable. The PROCESS macro for SPSS developed by Hayes (2012) was used. This macro uses an ordinary least squares analytic framework and estimates the conditional effects of the predictor when the moderator is equal to the mean, one standard deviation above the mean and one standard deviation below the mean. Predictor variables were mean-centred before the analysis.

Leadership and its predictors

Leadership, the RSES and D-IRAP. There were no significant main effects of the RSES ($b = .16$, $t = .74$, $p = .46$) or D-IRAP ($b = 4.49$, $t = 1.17$, $p = .25$) predicting Leadership. However, a significant interaction between the RSES and D-IRAP was found ($b = 2.9$, $t = 3.63$, $p < .001$). The model was significant ($F(3, 45) = 4.51$, $p < .001$)

and explained 28% of Leadership variance (see Fig. 1). The analysis of the conditional effects of the RSES on Leadership at values of the moderator showed that when D-IRAP was relatively low (-1 *SD*) there was a marginally significant negative relationship between the RSES and Leadership ($b = -.55$, $t = -1.95$, $p = .06$). At the mean value of D-IRAP no relationship was observed between the RSES and Leadership ($b = .16$, $t = .74$, $p = .46$). When D-IRAP was relatively high (+1 *SD*) there was a significant positive relationship between the RSES and Leadership ($b = .87$, $t = 2.92$, $p < .01$).

Leadership, the RSES and D-IRAP-POS. The model was not significant ($F(3,45) = .58$, $p = .63$) and there were neither significant main effects of the RSES ($b = .23$, $t = .83$, $p = .40$) or D-IRAP-POS ($b = 1.75$, $t = .51$, $p = .61$) nor their interaction effect ($b = .91$, $t = 1.06$, $p = .29$) on Leadership.

Leadership, the RSES and D-IRAP-NEG. There were no significant main effects of the RSES ($b = .15$, $t = .69$, $p = .49$) or D-IRAP-NEG ($b = 3.44$, $t = 1.35$, $p = .18$), but a significant interaction between the RSES and D-IRAP-NEG was found ($b = 2.43$, $t = 5.32$, $p < .001$). The model was significant ($F(3,45) = 9.51$, $p < .001$) and explained 34% of Leadership variance. The exploration of the conditional effects of the RSES on Leadership at values of the moderator showed the effects similar to those described above (for D-IRAP used as the moderator).

Demand for Admiration and its predictors

For Demand for Admiration (DA) there were two significant negative main effects of the RSES ($b = -.39$, $t = -2.54$, $p = .01$), and D-IRAP ($b = -7.08$, $t = -2.54$, $p = .03$) as well as a significant interaction between these variables ($b = 1.74$, $t = 2.54$, $p = .01$). The model was significant ($F(3,45) = 5.43$, $p < .01$) and explained 30% of DA variance (see Fig. 2). The analysis of the conditional

Figure 1. Predicted values for Leadership illustrating the interaction of explicit self-esteem and implicit self-esteem (D-IRAP) at one standard deviation above and below the mean
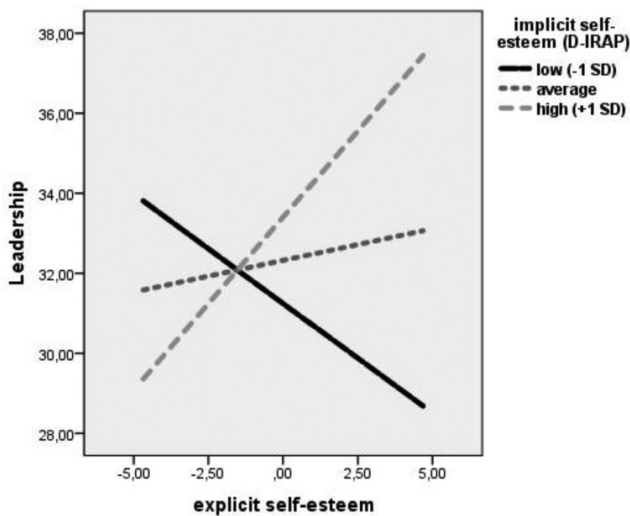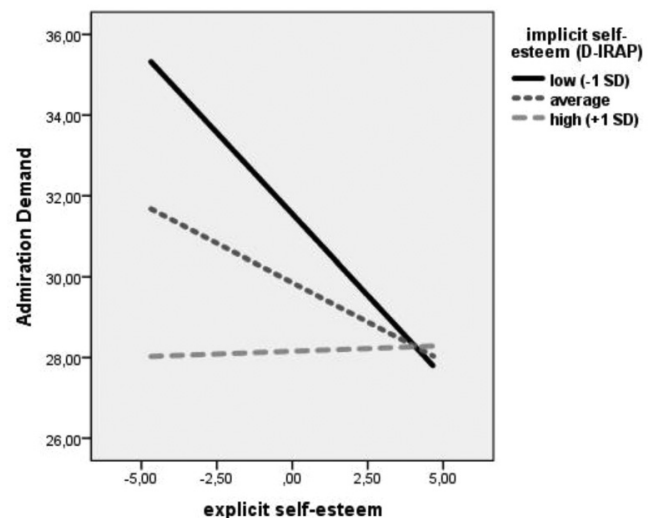

Figure 2. Predicted values for Demand for Admiration illustrating the interaction of explicit self-esteem and implicit self-esteem (D-IRAP) at one standard deviation above and below the mean

effects of the RSES on DA at values of the moderator showed that when D-IRAP was relatively low (-1 *SD*) there was a significant negative relationship between the RSES and DA ($b = -.80$, $t = -3.13$, $p < .01$). At the mean value of D-IRAP there was a weak negative relationship between the RSES and DA ($b = -.39$, $t = -2.54$, $p < .05$). When D-IRAP was relatively high (+1 *SD*) there was no relationship between the RSES and DA ($b = .03$, $t = .15$, $p = .88$). Analogical relationships were observed when D-IRAP-POS and D-IRAP-NEG were introduced as moderators.

Vanity and its predictors

There were neither significant main effects of the RSES or D-IRAP, nor interaction effect of these variables on Vanity. An identical result was obtained when D-IRAP-POS and D-IRAP-NEG were used as moderators. Thus, additional analyses were performed in which D-trial type 1, 2, 3 and 4 were used as moderators. A significant interaction between the RSES and Vanity was observed only for Self-Positive ($b = -.81$, $t = -1.99$, $p = .05$). The main effect of the RSES was marginally significant ($b = .19$, $t = 1.64$, $p = .10$), but there was no main effect of Self-Positive ($b = .48$, $t = .27$, $p = .79$). The entire model was non-significant ($F(3, 14) = 1.80$, $p = .16$) (see Fig. 3). The exploration of the conditional effects of the RSES on Vanity at values of the moderator indicated that when Self-Positive was relatively low (-1 *SD*) there was a significant positive relationship between the RSES and Vanity ($b = .45$, $t = 2.29$, $p = .03$). At the mean value of Self-Positive a marginally significant positive relationship was observed between the RSES and Vanity ($b = .19$, $t = 1.64$, $p = .10$). When Self-Positive was relatively high (+1 *SD*) there was no significant relationship between the RSES and Vanity ($b = .32$, $t = -.47$, $p = .64$).

Self-sufficiency and its predictors

Self-sufficiency, the RSES and D-IRAP. There was a significant positive main effect of the RSES on Self-sufficiency (S-s) ($b = .37$, $t = 4.16$, $p < .001$) and no significant main effect of D-IRAP ($b = -2.97$, $t = -1.53$, $p = .13$). However, a marginally significant interaction between the RSES and D-IRAP was observed ($b = .76$, $t = 1.84$, $p = .07$). The model was significant ($F(3,45) = 7.48$, $p < .001$) and explained 36% of S-s variance. The analysis of the conditional effects of the RSES on S-s at values of the moderator indicated that when D-IRAP was relatively low (-1 *SD*) there was no significant relationship between the RSES and S-s ($b = .1921$, $t = 1.51$, $p = .14$). A significant positive relationship was noted between the RSES and S-s at the mean value of D-IRAP ($b = .37$, $t = 4.16$, $p < .001$) and when D-IRAP was relatively high (+1 *SD*; $b = .55$, $t = 3.99$, $p < .001$).

Self-sufficiency, the RSES and D-IRAP-POS. Two main effects on S-s were found i.e. a significant positive main effect of the RSES ($b = .35$, $t = 3.27$, $p < .01$) and a significant negative main effect of D-IRAP-POS ($b = -3.03$, $t = -2.11$, $p = .04$). There was no interaction between the predictors ($b = -.05$, $t = -.17$, $p = .87$). The model was significant ($F(3,45) = 8.28$, $p < .001$) and explained 30% of S-s variance.

Self-sufficiency, the RSES and D-IRAP-NEG. Only one main significant positive effect of the RSES on S-s was observed ($b = .38$, $t = 4.68$, $p < .001$). The effect of D-IRAP-NEG was non-significant ($b = -1.06$, $t = -.80$, $p = .43$). However, there was a significant interaction between the RSES and D-IRAP-NEG in predicting S-s ($b = .91$, $t = 4.25$, $p < .001$). The model was significant ($F(3, 45) = 14.60$, $p < .001$) and explained 42% of S-s variance (see Fig. 4). The results of the analysis of the conditional effects were similar to those described above (for D-IRAP used as the moderator).

Figure 3. Predicted values for Vanity illustrating the interaction of explicit self-esteem and implicit self-esteem (D-IRAP-tt1) at one standard deviation above and below the mean
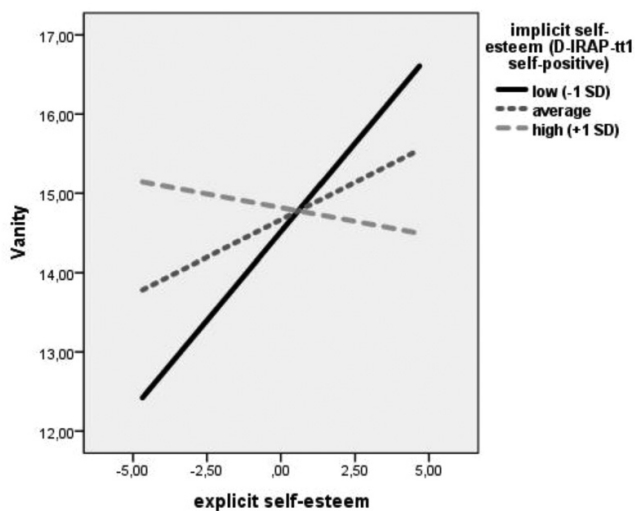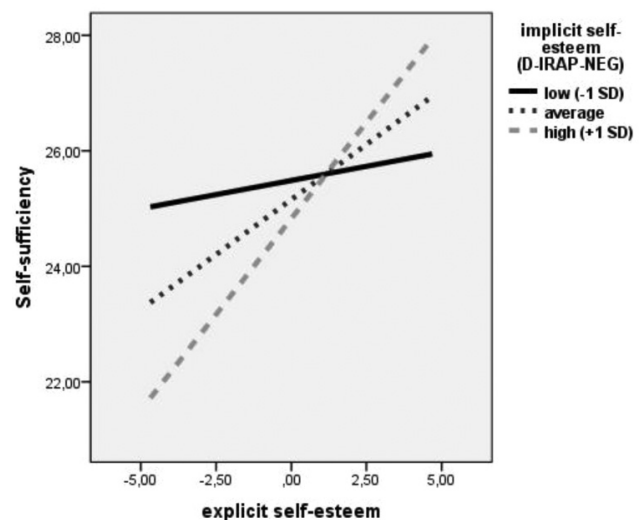

Figure 4. Predicted values for Self-sufficiency illustrating the interaction of explicit self-esteem and implicit self-esteem (D-IRAP-NEG) at one standard deviation above and below the mean

Narcissism and its predictors

There were no significant main effects of the RSES ($b = .31$, $t = .79$, $p = .43$) or D-IRAP ($b = -4.66$, $t = -.57$, $p = .57$) on narcissism. An interaction between the predictors was significant ($b = 5.22$, $t = 3.29$, $p < .01$) as well as the entire model ($F(3,45) = 4.48$, $p < .01$, 22% of variance explained). The analysis of the conditional effects of the RSES on narcissism showed a significant positive relationship between the predictor and narcissism ($b = 1.56$, $t = 3.25$, $p < .01$) only when D-IRAP was relatively high ($+1$ *SD*). When D-IRAP-NEG was a moderator, the relationship between narcissism and the RSES was similar. The analysis with D-IRAP-POS as the moderator did not reveal any relationships between the variables.

Discussion

The aim of the present study was the assessment of relationships between facets of grandiose narcissism and ESE/ISE. Our purpose was to compare the results with the three models presented by Gregg and Sedikides (2010). Additionally, we aimed at usefulness of a relatively new latency-based measure of ISE. The analysis concerning trial type scores showed that positivity bias, typical of ESE but also frequently observed in ISE assessed using various measures (Bosson et al., 2000; Greenwald & Farnham, 2000; Gregg & Sedikides, 2010; Timko, England, Herbert, & Forman, 2010), was present for three trial type scores (Self-Positive, Self-not Positive, Self-not Negative) but was not present in the case of self-negative scores (when the participants responded consistently with the rule assuming the possession of negative traits). In other words, responses of the participants, regardless of their level of narcissism, do not indicate a clear rejection of negative traits ascribed to the self. Such a result constitutes

an incentive for a separate analysis of scores for positive and negative words in accordance with the bidimensional conceptualisation of SE (Elliot & Mapes, 2005).

In the current study only one of the trial type scores (self-negative) showed a weak positive relationship with ESE. Interestingly, this correlation concerns the only trial type score in relation to which positivity bias was not observed. Generally, such a result is consistent with the results of the previous studies using other instruments for ISE measurement (Koole & DeHart, 2007). ESE and ISE measured using the IRAP were also only marginally related (see also Vahey, Barnes-Holmes, & Barnes-Holmes, 2009).

Correlations between ESE and narcissism were weaker compared to those that are usually obtained. The negative correlation between ESE and Demand for Admiration (subscale which in the Polish version of the NPI corresponds to the Entitlement, Superiority and Exhibitionism subscales developed by Raskin and Terry, 1988; see also Sorokowski, Sorokowska, Oleszkiewicz, Frackowiak, Huk, & Pisanski, 2015, p. 124) supports the opinion that this component of narcissism can reflect its potentially maladaptive elements. This result is consistent both with the results of the previous studies, in which a negative correlation between ESE and NPI subscales was observed (usually for Exploitativeness/Entitlement; e.g. Ackerman et al., 2011), and with the observations that grandiose narcissism contains some elements of vulnerability, which leads to opposing relationships between NPI subscales and other variables (e.g. Ruiz, Smith, & Rhodewald, 2001). Due to the negative relationship between DA and ESE, a positive correlation between the NPI total score and ESE became non-significant in this sample. Moreover, low positive correlations between the three NPI subscales and the RSES (non-significant for Leadership and Vanity) are consistent

with the observations that correlation between narcissism and ESE is lower when the items included in the SE questionnaire concern domination-related situations only to a small extent (since the RSES is considered that type of measure; Brown & Zeigler-Hill, 2004).

Relationships between ESE, ISE and Leadership

The Leadership subscale reflects the beliefs concerning one's own leadership skills and ability to exert influence on others. Regression analysis indicated that ISE and ESE were not significant predictors of Leadership. However, crossover interaction occurred between these two types of SE when D-IRAP or D-IRAP-NEG was used. ISE moderated the relationship between ESE and Leadership in such a way that for low ISE the correlation between ESE and Leadership was negative, whereas for high ISE this correlation was positive. There was no relationship between ESE and Leadership in individuals with average ISE. The expected positive relationship between ESE and Leadership occurred only in individuals whose ISE was relatively high. The highest level of this type of narcissism was observed in individuals with **congruent high SE** which is opposed to the models which assume that discrepant SE is the basis of narcissism.

While interpreting this unexpected result, it may be assumed that the Leadership subscale does not measure narcissism, but rather reflects beliefs of a confident, go-getting individual about having leadership competence. However, the Leadership subscale consists also of items which describe the beliefs related to manipulation, which in the original version of the NPI are included in the Exploitativeness/Entitlement factor, and in the validation studies moderately correlated with Demand for Admiration and Vanity (Bazińska & Drat-Ruszczak, 2000). It is possible that in the case of this facet of narcissism, the uncertainty of one's SE ascribed to narcissists is manifested in a different way, e.g. as unstable SE. It is also possible that Leadership represents a separate type of narcissism unrelated to fragile ego. In accordance with the results of the study which aimed at the analysis of the NPI factor structure (Ackerman et al., 2011), Leadership was the only subscale (out of the three subscales) which was related to adaptive outcomes and negatively correlated with contingent SE. Similarly, Zeigler-Hill and Besser (2013, p. 258) found that "the facet of narcissism that concerns leadership appears to preserve the self-esteem of individuals in the face of social rejection or exclusion".

Relationships between ESE, ISE and Demand for Admiration

The DA subscale, which reflects the need to be important, admired and complimented, is the only dimension of narcissism which correlated negatively with ESE and ISE. A negative correlation between DA and SE suggests that DA diagnoses **vulnerable** type of narcissism. This type of results (only for ESE) were previously obtained in the studies using the original version of the NPI (e.g. Brown, Budzek, & Tamborski,

2009). Additionally, ISE moderated the relationship between ESE and DA – together with an increase in ISE, the negative relationship between ESE and DA decreased and became non-significant. The highest level of DA was noted in individuals with **congruent low SE**, whereas individuals with discrepant high, discrepant low and congruent high SE demonstrated a low level of this trait. Such a result is not consistent with any of the models describing relationships between SE and narcissism (Gregg & Sedikides, 2010). However, it remains consistent with the results of the study of Conner and Barrett (2005), concerning the relationships between ISE/ESE and experiencing emotions in everyday life, which indicated a particular disposition of individuals with congruent low SE to experience negative affect. In this study individuals with discrepant high (fragile), discrepant low (defensive) and congruent high SE did not differ significantly and demonstrated a low level of negative affect. Such patterns of ESE and ISE interactions predicting emotional vulnerability are fully consistent with the results for DA obtained in the present study but are inconsistent with the results of a number of other studies on SE (Bosson et al., 2003; Jordan et al., 2003; Kernis, 2003). One of the major traits of individuals with vulnerable type of narcissism is the disposition to experience negative emotions (stress and low well-being). Consequently, the results of the study by Conner and Barrett (2005) indirectly support the possibility that congruent low SE can be the basis of this type of narcissism. The fact that individuals with discrepant low SE achieved relatively low DA scores can be regarded as consistent with the idea that in individuals with low ESE high ISE can have a positive function. In this case it is manifested as weaker inclination to display grandiose tendencies and seek acclaim. This result is thus consistent with the "glimmer of hope" hypothesis and demonstrates that discrepant low SE does not always have to be the source of vulnerability (Spencer et al., 2005; Jordan et al., 2013).

Relationships between ESE, ISE and Vanity

Vanity is the NPI subscale which reflects "love towards oneself" and approval of one's own appearance. Neither ESE/ISE nor the interaction between them proved to be significant predictors of Vanity. An additional analysis in which D-IRAP-tt1 (self-positive) was used revealed significant crossover interaction between ESE and ISE: the RSES correlated positively with Vanity only when D-IRAP-tt1 scores were lower. The highest level of Vanity was observed in individuals with high ESE and low ISE, which is consistent with the **partial discrepancy (mask) model**. The association of a high level of Vanity with discrepant high SE, treated as a possible cause of doubt about self-worth which torments a narcissistic (vain) individual, was visible only for one ISE indicator (related to the task of confirming one's own positive traits). In other words, in accordance with the above-mentioned result, vain individuals at the unconscious level can be uncertain whether they possess desirable traits (whether

they are sufficiently perfect). However, they do not exhibit uncertainty (greater than the others) related to possessing negative traits. High level of Vanity was noted in these participants with high SE who nurture implicit uncertainty about their own virtues. If lack of uncertainty was observed, there was no relationship between tendency towards vanity and ESE.

Relationships between ESE, ISE and Self-sufficiency

Self-sufficiency is a dimension of narcissism related to the conviction about one's own independence, individualism, reliability and competence. In the present study, it correlated positively with ESE and negatively with ISE (only for D-IRAP-POS). The results of the moderation analysis for this NPI subscale were different depending on which ISE indicator was taken into consideration. If D-IRAP-POS was used, two types of SE separately determined S-s (ESE positively and ISE negatively) and they did not interact with each other. The highest level of S-s was observed in individuals with high ESE and low ISE. This result is consistent with the **global marker model**. However, when D-IRAP-NEG was used, only ESE predicted S-s, whereas ISE functioned as the moderator of this relation. That means that this relationship was strong in individuals with high ISE, then it gradually declined and eventually became non-significant in individuals with low ISE. The highest level of S-s was observed in participants with **congruent high SE**, whereas the lowest in those with discrepant low SE. In participants with low ESE, the level of S-s was diverse and the relatively highest in participants with the lowest ISE (congruent low SE). Such a result is not consistent with the predictions of the previously described models. However, it is reasonable to believe that for individuals with congruent low SE who do not have a protective "buffer" created by positive self-view, assuming the attitude of independence and autonomy can be the most accessible and, at the same time, an effective way of protection (see also Horney, 1945).

The way of ISE measurement (using positive or negative traits) differentiated the relationships between SE and S-s. Considering the responses of the participants that were based on accepting or denying positive traits, the results supported the global marker model, according to which ESE and ISE are separate predictors of narcissism. As in the case of Vanity, a high level of S-s was observed in the individuals who at the unconscious level are uncertain about possession of positive traits. It does not apply to negative traits though. Considering responses that were based on accepting or denying negative traits, the relationships between ESE and ISE in predicting S-s were more complex. The highest level of S-s was noted in individuals with congruent high (secure) SE, which is regarded as the most beneficial for an individual. It can be speculated that individuals with such a pattern of relationships between ESE and ISE exhibit self-confidence and a sense of identity expressing strong and stable self. High ISE in these individuals reflects not so much the feeling of one's own perfection (positive traits), but rather lack of association between the self and negative traits.

Relationships between ESE, ISE and narcissism

ESE and ISE were not predictors of narcissism (total score) in the regression analysis, although crossover interaction between two types of SE was noted (for D-IRAP and D-IRAP-NEG). A positive correlation between ESE and narcissism occurred if ISE was high and negative correlation was observed if ISE was low. The highest level of narcissism was demonstrated by individuals with congruent high SE and congruent low SE. When D-IRAP-POS was the moderator of the relations between ESE and narcissism, the model was non-significant. According to these results, which are not consistent with any of the models describing configurations of SE in narcissists, individuals with congruent high and congruent low SE were characterised by a high level of narcissism. However, the comparison of the results of the described analyses conducted at the level of facets of narcissism indicates that each of these components of narcissism displayed a different pattern of relations with SE, additionally differentiated due to the type of the IRAP score which was used as the moderator. Their combination in the composite score leads not only to loss of information, but can also pose difficulties in interpretation (Ruiz et al., 2001; Ackerman et al., 2011; Brown et al., 2009).

Conclusions

The current study is the first which applies the IRAP for the analysis of relationships between ESE/ISE and narcissism. To the best of our knowledge, this is also the first investigation of the relationships between the particular facets of grandiose narcissism and ESE/ISE.

It is possible that there is no simple answer to the question of ESE and ISE configuration directly related to narcissism, as it may depend both on the type of narcissism and the type of SE. Grandiose narcissism and vulnerable narcissism (not measured in the current study) constitute separate constructs which differ from each other by their relationship to ESE. Additionally, separate elements can also be distinguished within grandiose narcissism. In this study they are represented by the remaining three subscales of the NPI. Different relationships between these narcissism components and ESE/ISE observed in the current study may reflect various ways leading to their construction in the process of development, currently described by competitive concepts (Morf & Rhodewalt, 2001; Millon, 1981). Therefore, the present analysis supports the reports on NPI heterogenity and is consistent with the recent reports (Ackerman et al., 2011; Falkenbach, Howe, & Falki, 2013) confirming that NPI allows to measure two types of narcissism (i.e. more or less "healthy" one).

The results of the present study provided limited support for the two models mentioned by Gregg and Sedikides (2010) describing the relations between narcissism and SE. The global marker model was supported in the case of Self-sufficiency (only when ISE was measured using positive words). The relationships between Vanity and SE were consistent with partial discrepancy (mask) model (only for self-positive D-IRAP score). In the case of Leadership, none of the models were applicable (the highest level of this trait was observed in individuals with congruent

high SE), thus narcissistic fragile self can only be found in contingent and unstable SE. It is also possible that this dimension of narcissism is not related to uncertainty or self-doubts (Bosson et al., 2008) but rather reflects a potentially healthy form of narcissism. On the other hand, Demand for Admiration seemed to capture vulnerable type of narcissism as it was negatively related to both ESE and ISE.

SE can be studied as a simple reflection of global self-attitude or as a complex construct. The IRAP seems to be an instrument properly adjusted to a more in-depth study of ISE, despite limited reliability and validity, which is a common weakness of all implicit measures. Positive D-IRAP score can include both elements (accepting positive traits and denying negative traits) equally or can be based more on one of them. According to the recommendations, the overall D-IRAP scores may be calculated (Hussey et al., 2015). However, then the benefits of the IRAP application are lost as it is an instrument more complex and it places a larger burden on participants compared to the IAT. In the analysis consideration of particular IRAP scores (based on responses in each trial type and on responses to positive or negative traits) appears to be particularly interesting in reference to narcissism. It allows to establish the type of uncertainty related to the self which may be shared by narcissistic individuals. Narcissists are believed to have a particularly strong tendency to create separate representations of "good" and "bad" self (Zeigler-Hill, Myers, & Clark, 2010), possibly with manifestations at the unconscious level. In the current study, these two types of ISE (concerning positive and negative traits) entered into different relationships with particular types of narcissism.

In the current study, various patterns of relationships between ESE/ISE and the particular facets of narcissism were obtained. The above-mentioned differences concern the results for the NPI total score and subscale scores as well as relationships between predictors (in the case of the application of D-IRAP or its components). These differences indicate that composite measures which constitute a resultant of separate elements do not reflect the entire complexity of the analysed phenomenon in the best scenario. In the worst scenario these differences depict a false image of the phenomenon.

The described study has some limitations due to the fact that it was conducted on a relatively small sample of students. Therefore, the results should be interpreted cautiously. Target words used in the IRAP were mostly related to agentic traits which are valued and aspired to by narcissistic individuals. Repeating the study using communal words would be highly beneficial. The Polish version of NPI was used in this study, and its subscales do not map directly on the original NPI scale, which may hamper direct comparison of the results.

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*, 67–87.

Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a direct measure of implicit beliefs. *The Irish Psychologist, 32*, 169–177.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record, 60*, 527–542.

Barnes-Holmes, D., Murtagh, L., Barnes-Holmes, Y., & Stewart, I. (2010). Using the Implicit Association Test and the Implicit Relational Assessment Procedure to measure attitudes toward meat and vegetables in vegetarians and meat-eaters. *Psychological Record, 60*, 287–305.

Baumeister, R.F., Campbell, J.D., Krueger, J.I., & Vohs, K.D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest, 4*, 1–44.

Baumeister, R.F., Smart, L., & Boden, J.M. (1996). Relation of threatened egotism to violence and aggression: the dark side of high self-esteem. *Psychological Review, 103*, 5–33.

Bazińska, R., & Drat-Ruszczak, K. (2000). Struktura narcyzmu w polskiej adaptacji kwestionariusza NPI Raskina i Halla [Structure of narcissism in Polish adaptation of Raskin's and Hall's NPI]. *Czasopismo Psychologiczne, 6*, 171–188.

Besser, A., & Priel, B. (2009). Emotional Responses to a Romantic Partner's Imaginary Rejection: The Roles of Attachment Anxiety, Covert Narcissism, and Self-Evaluation. *Journal of Personality, 77*, 287–325.

Boduszek, D., Hyland, P., Dhingra, K., & Mallett, J. (2013). The Factor structure and composite reliability of the Rosenberg Self-Esteem Scale among ex-prisoners. *Personality and Individual Differences, 55*, 877–881.

Boldero, J.M., Higgins, E.T., & Hulbert, C.A. (2015). Self-regulatory and narcissistic grandiosity and vulnerability: Common and discriminant relations. *Personality and Individual Differences, 76*, 171–176.

Bosson, J., & Swann, W.B., Jr. (2009). Self-esteem: Nature, origins, and consequences. In R. Hoyle & M. Leary (Eds.), *Handbook of individual differences in social behavior* (pp. 527–546). New York: Guilford.

Bosson, J.K., & Prewitt-Freilino, J.L. (2007). Overvalued and ashamed: Considering the roles of self-esteem and self-conscious emotions in covert narcissism. In J.L. Tracy, R.W. Robins & J.P. Tangney (Eds.), *The Self-conscious Emotions: Theory and Research* (2nd ed., pp. 407–425). New York, NY: Guilford.

Bosson, J.K., Brown, R.P., Zeigler-Hill, V., & Swann, W.B., Jr. (2003). Self-enhancement tendencies among people with high explicit self-esteem: The moderating role of implicit self-esteem. *Self and Identity, 2*, 169–187.

Bosson, J.K., Lakey, C.E., Campbell, W.K., Zeigler-Hill, V., Jordan, C.H., & Kernis, M.H. (2008). Untangling the Links between Narcissism and Self-esteem: A Theoretical and Empirical Review. *Social and Personality Psychology Compass, 2*, 1415–1439.

Bosson, J.K., Swann Jr, W.B., & Pennebaker, J.W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79*, 631–643.

Brown, R.P., & Zeigler-Hill, V. (2004). Narcissism and the non-equivalence of self-esteem measures: A matter of dominance?. *Journal of Research in Personality, 38*, 585–592.

Brown, R.P., Budzek, K., & Tamborski, M. (2009). On the meaning and measure of narcissism. *Personality and Social Psychology Bulletin, 35*, 951–964.

Buhrmester, M.D., Blanton, H., & Swann, W.B., Jr. (2011). Implicit self-esteem: nature, measurement, and a new way forward. *Journal of Personality and Social Psychology, 100*, 365–385.

Cain, N.M., Pincus, A.L., & Ansell, E.B. (2008). Narcissism at the crossroads: Phenotypic description of pathological narcissism across clinical theory, social/personality psychology, and psychiatric diagnosis. *Clinical Psychology Review, 28*, 638–656.

Campbell, W.K., Bosson, J.K., Goheen, T.W., Lakey, C.E., & Kernis, M.H. (2007). Do narcissists dislike themselves "deep down inside"? *Psychological Science, 18*, 227–229.

Campbell, W.K., Rudich, E., & Sedikides, C. (2002). Narcissism, self-esteem, and the positivity of self-views: Two portraits of self-love. *Personality and Social Psychology Bulletin, 28*, 358–368.

Conner, T., & Barrett, L.F. (2005). Implicit Self-Attitudes Predict Spontaneous Affect in Daily Life. *Emotion*, *5*, 476–488.

del Rosario, P.M., & White, R.M. (2005). The Narcissistic Personality Inventory: Test–retest stability and internal consistency. *Personality and Individual Differences*, *39*, 1075–1081.

Dickinson, K.A., & Pincus, A.L. (2003). Interpersonal analysis of grandiose and vulnerable narcissism. *Journal of Personality Disorders*, *17*, 188–207.

Dijksterhuis, A., Albers, L., & Bongers, K. (2008). Digging for the real attitude: Lessons from research on implicit and explicit self-esteem. In R.E. Petty, R.H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new wave of implicit measures* (pp. 229–250). New York: Psychology Press.

Drake, C.E., Kellum, K.K., Wilson, K.G., Luoma, J.B., Weinstein, J.H., & Adams, C.H. (2011). Examining the implicit relational assessment procedure: Four preliminary studies. *The Psychological Record*, *60*, 81–100.

Elliot, A.J., Mapes, R.R. (2005). Approach-avoidance motivation and self-concept evaluation. In A. Tesser, J.V. Vood, D.A. Stapel, *On building, defending, and regulating the self: A psychological perspective* (pp. 171–196). New York: Psychology Press.

Emmonds, R.A. (1984). Factor analysis and construct validity of the Narcissistic Personality Inventory. *Journal of Personality Assessment 48*, 291–300.

Epstein, S. (2006). Conscious and unconscious self-esteem from the perspective of cognitive-experiential self-theory. In M.H. Kernis (Ed.), *Self-esteem issues and answers: A sourcebook of current perspectives* (pp. 69–76). New York, NY: Psychology Press.

Falkenbach, D.M., Howe, J.R., & Falki, M. (2013). Using self-esteem to disaggregate psychopathy, narcissism, and aggression. *Personality and Individual Differences*, *54*, 815–820.

Farwell, L., & Wohlwend-Lloyd, R. (1998). Narcissistic processes: Optimistic expectations, favorable self-evaluations, and self-enhancing attributions. *Journal of Personality, 66*, 65–83.

Fazio, R.H., & Olson, M.A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54,* 297–327.

Franck, E., De Raedt, R., & De Houwer, J. (2007). Implicit but not explicit self-esteem predicts future depressive symptomatology. *Behaviour Research and Therapy*, *45*, 2448–2455.

Given-Wilson, Z., McIlwain, D., & Warburton, W. (2011). Meta-cognitive and interpersonal difficulties in overt and covert narcissism. *Personality and Individual Differences*, *50*, 1000–1005.

Golijani-Moghaddam, N., Hart, A., & Dawson, D.L. (2013). The implicit relational assessment procedure: Emerging reliability and validity data. *Journal of Contextual Behavioral Science*, *2*, 105–119.

Greenwald, A.G., & Farnham, S.D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*, 1022–1038.

Gregg, A.P., & Sedikides, C. (2010). Narcissistic fragility: Rethinking its links to explicit and implicit self-esteem. *Self and Identity*, *9*, 142–161.

Grumm, M., Nestler, S., & Von Collani, G. (2009). Changing explicit and implicit attitudes: The case of self-esteem. *Journal of Experimental Social Psychology*, *45*, 327–335.

Hayes, A.F. (2012). *PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling* [White paper]. Retrieved from http://www.afhayes.com/public/process2012.pdf

Hendin, H.M., & Cheek, J.M. (1997). Assessing hypersensitive narcissism: A reexamination of Murray's Narcissism Scale. *Journal of Research in Personality*, *31*, 588–599.

Horvath, S., & Morf, C.C. (2009). Narcissistic defensiveness: Hypervigilance and avoidance of worthlessness. *Journal of Experimental Social Psychology*, *45*, 1252–1258.

Hyland, P., Boduszek, D., Dhingra, K., & Shevlin, M. (2014). A Bifactor Approach to Modelling the Rosenberg Self-Esteem Scale. *Personality and Individual Differences*, *66*, 188–192.

Hofmann, W., Gawronski, G., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31,* 1369–1385.

Horney, K. (1945). *Our Inner conflicts*. New York: Norton.

Horton, R.S., Bleau, G., & Drwecki, B. (2006). Parenting narcissus: What are the links between parenting and narcissism? *Journal of Personality*, *74*, 345–376.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorizing in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record*, *61*, 465–496.

Hussey, I., & Barnes-Holmes, D. (2012). The Implicit Relational Assessment Procedure as a measure of implicit depression and the role of psychological flexibility. *Cognitive and Behavioral Practice, 19*, 573–582.

Hussey, I., Thompson, M., McEnteggart, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2015). Interpreting and inverting with less cursing: A guide to interpreting IRAP data. *Journal of Contextual Behavioral Science*.

Jordan, C.H., Logel, C., Spencer, S.J., Zanna, M.P., Wood, J.V., & Holmes, J.G. (2013). Responsive low self-esteem: Low explicit self-esteem, implicit self-esteem, and reactions to performance outcomes. *Journal of Social and Clinical Psychology*, *32*, 703–732.

Jordan, C.H., Spencer, S.J., Zanna, M.P., Hoshino-Browne, E., & Correll, J. (2003). Secure and defensive high self-esteem. *Journal of Personality and Social Psychology, 85*, 969–978.

Jordan, C.H., Whitfield, M., & Zeigler-Hill, V. (2007). Intuition and the correspondence between implicit and explicit self-esteem. *Journal of Personality and Social Psychology*, *93*, 1067–1079.

Karpinski, A., & Steinman, R.B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*, 16–32.

Kernberg, O.F. (1970). Factors in the psychoanalytic treatment of narcissistic personalities. *Journal of the American Psychoanalytic Association, 18*, 51–85.

Kernis, M.H. (2003). Toward a conceptualization of optimal self-esteem. *Psychological Inquiry, 14*, 1–26.

Kernis, M.H., Abend, T., Shrira, I., Goldman, B.M., Paradise, A., & Hampton, C. (2005). Self-serving responses arising from discrepancies between explicit and implicit self-esteem. *Self and Identity*, *4*, 311–330.

Klavina, E., Schröder-Abé, M., & Schütz, A. (2012). Facets of self-esteem at an implicit level? Investigation of implicit–explicit correlations and development of four IATs. *Personality and Individual Differences*, *53*, 693–698.

Kohut, H. (1976). *The restoration of the self*. New York: International Universities Press.

Koole, S.L., & DeHart, T. (2007). Self-affection without self-reflection: Origins, models, and consequences of implicit self-esteem. In C. Sedikides & S. Spencer (Eds.), *The self in social psychology* (pp. 36–86). New York: Psychology Press.

Koole, S.L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology, 80*, 669–685.

Kosnes, L., Whelan, R., O'Donovan, A., & McHugh, L.A. (2013). Implicit measurement of positive and negative future thinking as a predictor of depressive symptoms and hopelessness. *Consciousness and Cognition*, *22*, 898–912.

Krizan, Z., & Suls, J. (2008). Are implicit and explicit measures of self-esteem related? A meta-analysis for the Name–Letter Test. *Personality and Individual Differences, 44*, 521–531.

Łaguna, M., Lachowicz-Tabaczek, K., & Dzwonkowska, I. (2007). Skala samooceny SES Morrisa Rosenberga – polska adaptacja metody [Morris Rosenberg's self-esteem scale SES – a Polish adaptation of the method]. *Psychologia Społeczna, 2*, 164–176.

Lannin, D.G., Guyll, M., Krizan, Z., Madon, S., & Cornish, M. (2014). When are grandiose and vulnerable narcissists least helpful? *Personality and Individual Differences*, *56*, 127–132.

Luchner, A.F., Houston, J.M., Walker, C., & Houston, M.A. (2011). Exploring the relationship between two forms of narcissism and competitiveness. *Personality and Individual Differences*, *51*, 779–782.

Maxwell, K., Donnellan, M.B., Hopwood, C.J., & Ackerman, R.A. (2011). The two faces of Narcissus? An empirical comparison of the Narcissistic Personality Inventory and the Pathological Narcissism Inventory. *Personality and Individual Differences, 50*, 577–582.
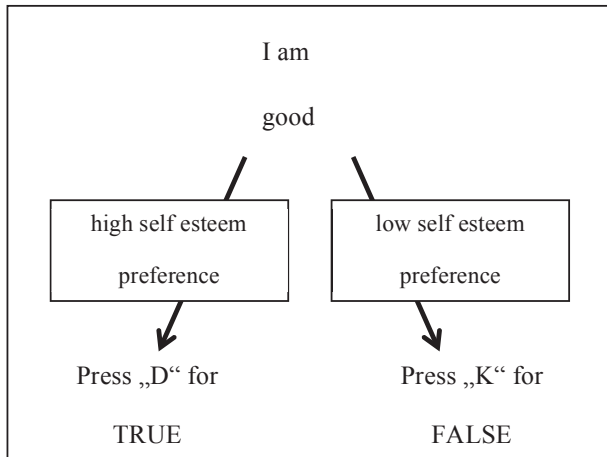
McKay, M., Boduszek, D., & Harvey, S. (2014). The Rosenberg Self-Esteem Scale: A bifactor answer to two-factor question? *Journal of Personality Assessment*, *96*, 654–660.

Miller, J.D., & Campbell, W.K. (2008). Comparing Clinical and Social-Personality Conceptualizations of Narcissism. *Journal of Personality*, *76*, 449–476.

Miller, J.D., Dir, A., Gentile, B., Wilson, L., Pryor, L.R., & Campbell, W.K. (2010). Searching for a vulnerable dark triad: Comparing factor 2 psychopathy, vulnerable narcissism, and borderline personality disorder. *Journal of Personality, 78*, 1529–1564.

Miller, J.D., Gentile, B., Wilson, L., & Campbell, W.K. (2013). Grandiose and vulnerable narcissism and the DSM–5 pathological personality trait model. *Journal of Personality Assessment*, *95*, 284–290.

Millon, T. (1981). *Disorders of Personality*. New York, NY: Wiley.

Morf, C.C., & Rhodewalt, F. (2001). Unraveling the paradoxes of narcissism: A dynamic self-regulatory processing model. *Psychological Inquiry, 12*, 177–196.

Nicholson, E., & Barnes-Holmes, D. (2012b). The Implicit Relational Assessment Procedure (IRAP) as a measure of spider fear. *The Psychological Record, 62*, 263–277.

Nosek, B.A., & Banaji, M.R. (2001). The go/no-go association task. *Social Cognition*, *19*, 625–666.

Nosek, B.A., Hawkins, C.B., & Frazier, R.S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences*, *15*, 152–159.

Olson, M.A., & Fazio, R.H. (2008). Implicit and explicit measures of attitudes: The perspective of the MODE model. In R.E. Petty, R.H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new wave of implicit measures* (pp. 19–65). New York: Psychology Press.

Pelham, B.W., Koole, S.L., Hardin, C.D., Hetts, J.J., Seah, E., & DeHart, T. (2005). Gender moderates the relation between implicit and explicit self-esteem. *Journal of Experimental Social Psychology*, *41*, 84–89.

Pincus, A.L., Ansell, E.B., Pimentel, C.A., Cain, N.M., Wright, A.G., & Levy, K.N. (2009). Initial construction and validation of the Pathological Narcissism Inventory. *Psychological Assessment, 21*, 365–379.

Raskin, R.N., & Hall, C.S. (1979). A narcissistic personality inventory. *Psychological Reports*, *45*, 590.

Raskin, R.N., & Hall, C.S. (1981). The Narcissistic Personality Inventory: Alternative form reliability and further evidence of construct validity. *Journal of Personality Assessment*, *45*, 159–162.

Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology, 54*, 890–902.

Rose, P. (2002). The happy and unhappy faces of narcissism. *Personality and Individual Differences*, *33*, 379–391.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rudman, L.A., Dohn, M.C., & Fairchild, K. (2007). Implicit self-esteem compensation: automatic threat defense. *Journal of Personality and Social Psychology*, *93*, 798–813.

Rudolph, A., Schröder-Abé, M., Riketta, M., & Schütz, A. (2010). Easier When Done Than Said!. *Zeitschrift für Psychologie/Journal of Psychology*, *218*, 12–19.

Ruiz, J.M., Smith, T.W., & Rhodewalt, F. (2001). Distinguishing narcissism and hostility: Similarities and differences in interpersonal circumplex and five-factor correlates. *Journal of Personality Assessment, 76*, 537–555.

Schreiber, F., Bohn, C., Aderka, I.M., Stangier, U., & Steil, R. (2012). Discrepancies between implicit and explicit self-esteem among adolescents with social anxiety disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*, 1074–1081.

Schröder-Abé, M., Rudolph, A., & Schütz, A. (2007). High implicit self-esteem is not necessarily advantageous: Discrepancies between explicit and implicit self-esteem and their relationship with anger expression and psychological health. *European Journal of Personality*, *21*, 319–339.

Schröder-Abé, M., Rudolph, A., Wiesner, A., & Schütz, A. (2007). Self--esteem discrepancies and defensive reactions to social feedback. *International Journal of Psychology*, *42*, 174–183.

Sedikides, C., & Gregg, A.P. (2001). Narcissists and feedback: Motivational surfeits and motivational deficits. *Psychological Inquiry, 12*, 237–239.

Sedikides, C., Rudich, E.A., Gregg, A.P., Kumashiro, M., & Rusbult, C. (2004). Are normal narcissists psychologically healthy? Self-esteem matters. *Journal of Personality and Social Psychology, 87*, 400–416.

Sorokowski, P., Sorokowska, A., Oleszkiewicz, A., Frackowiak, T., Huk, A., & Pisanski, K. (2015). Selfie posting behaviors are associated with narcissism among men. *Personality and Individual Differences*, *85*, 123–127.

Spencer, S.J., Jordan, C.H., Logel, C., & Zanna, M.P. (2005). Nagging doubts and a glimmer of hope: The role of implicit self-esteem in self-image maintenance. In A. Tesser, J.V. Wood, & D.A. Stapel (Eds.), *On building, defending, and regulating the self: A psychological perspective* (pp. 153–170). Philadelphia: Psychology Press.

Tennen, H., & Affleck, G. (1993). The puzzles of self-esteem: A clinical perspective. In R.F. Baumeister (Ed.), *Self-esteem: The puzzle of low self-regard* (pp. 241–262). New York: Plenum Press.

Timko, C.A., England, E.L., Herbert, J.D., & Forman, E.M. (2010). The Implicit Relational Assessment Procedure as a measure of self-esteem. *The Psychological Record*, *60*, 679–698.

Vahey, N.A., Barnes-Holmes, D., & Barnes-Holmes, Y. (2009). A first test of the Implicit Relational Assessment Procedure (IRAP) as a measure of self-esteem: Irish prisoner groups and university students. *The Psychological Record*, *59*, 371–388.

Vahey, N.A., Nicholson, E., & Barnes-Holmes, D. (2015). A meta-analysis of criterion effects for the Implicit Relational Assessment Procedure (IRAP) in the clinical domain. *Journal of Behavior Therapy and Experimental Psychiatry*, *48*, 59–65.

Vater, A., Ritter, K., Schröder-Abé, M., Schütz, A., Lammers, C.H., Bosson, J. K., & Roepke, S. (2013). When grandiosity and vulnerability collide: Implicit and explicit self-esteem in patients with narcissistic personality disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *44*, 37–47.

Vater, A., Schröder-Abé, M., Schütz, A., Lammers, C.H., & Roepke, S. (2010). Discrepancies between explicit and implicit self-esteem are linked to symptom severity in borderline personality disorder. *Journal of Behavior Therapy and Experimental Psychiatry, 41*, 357–364.

Watson, P.J., Varnell, S.P., & Morris, R.J. (1999–2000). Self-reported narcissism and perfectionism: an ego-psychological perspective and the continuum hypothesis. *Imagination, Cognition and Personality, 19*, 59–69.

Wierzchon, M. (2009). Czego nie wiemy, a czego tylko wydaje nam się, że nie wiemy o sobie [What we do not know and what we think we do not know about ourselves] In A. Niedźwieńska i J. Neckar (eds.), *Poznaj samego siebie czyli o źródłach samowiedz*y [*Meet yourself – about the origin of self-knowledge*] (pp. 179–193). Warsaw: Academica.

Wink, P. (1991). Two faces of narcissism. *Journal of Personality and Social Psychology*, *61*, 590–597.

Yamaguchi, S., Greenwald, A.G., Banaji, M.R., Murakami, F., Chen, D., Shiomura, K., Kobayashi, C., Cai, H., & Krendl, A. (2007). Apparent universality of positive implicit self-esteem. *Psychological Science*, *18*, 498–500.

Zeigler-Hill, V. (2006). Discrepancies between implicit and explicit self--esteem: Implications for narcissism and self-esteem instability. *Journal of Personality*, *74*, 119–143.

Zeigler-Hill, V., & Besser, A. (2013). A glimpse behind the mask: facets of narcissism and feelings of self-worth. *Journal of Personality Assessment, 95*, 249–260.

Zeigler-Hill, V., Myers, E.M., & Clark, C.B. (2010). Narcissism and self-esteem reactivity: The role of negative achievement events. *Journal of Research in Personality*, *44*, 285–292.

Irena Pilch, Magdalena Hyla
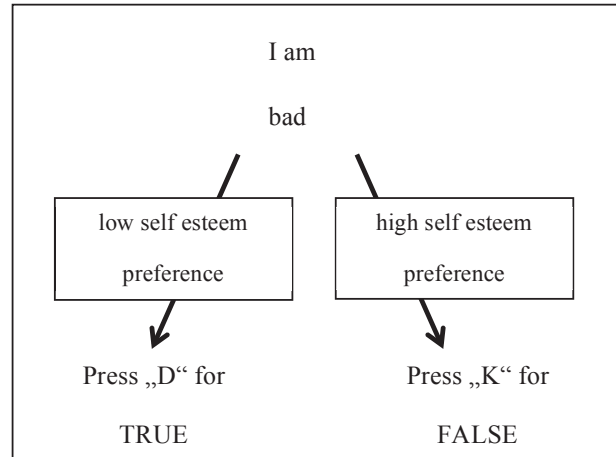
Appendix A

Figure A.1. Examples of the four trial types in the self-esteem IRAP

(Arrows with text boxes did not appear on-screen and are shown for illustrative purposes only)
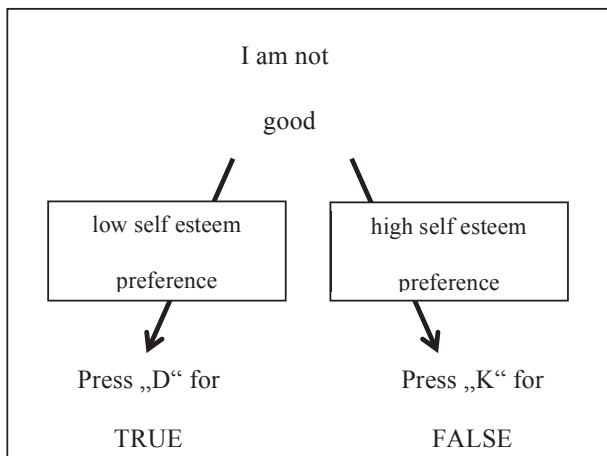
Trial type 1 - "I am positive"

I am

good

| high self esteem preference | low self esteem preference |

Press „D" for | Press „K" for

TRUE | FALSE

Trial type 2 = "I am negative"

I am

bad

| low self esteem preference | high self esteem preference |

Press „D" for | Press „K" for

TRUE | FALSE

Trial type 3 = "I am not positive"

I am not

good

| low self esteem preference | high self esteem preference |

Press „D" for | Press „K" for

TRUE | FALSE

Trial type 4 = "I am not negative"

I am not

bad

| high self esteem preference | low self esteem preference |

Press „D" for | Press „K" for

TRUE | FALSE