Erasmus
School of
Economics

ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

Bachelor Thesis

International Bachelor Economics and Business Economics

Personality and the prediction of entrepreneurial intentions

***Student*:** **E.S. olde Boerrigter** **492762**

Supervisor: **A.B. Wismans**

Second assessor: **S. Ramezani**

Date final version: **16-07-2021**

An individual's personality is often credited for its ability to guide people towards essential life choices, such as occupation. One such occupational choice is entrepreneurship, an essential factor in the economy. Current research on the relationship between personality and the intention to become an entrepreneur (EI) is inconsistent, instigating a debate on whether the "entrepreneurial personality" exists. This paper aims to add to the debate by using regression analysis to explore the association between EI and the personality traits of two models of personality: the Big Five Personality Traits and the Behavioural Inhibition and Behavioural Approach System (BIS/BAS). We then use these personality traits to predict EI using various classification methods (Logistic Regression, Naïve Bayes, and Random Forest). A moderation analysis on the role of gender in the relationship between EI and personality is also held. The analysis is conducted using a dataset on students at Erasmus University Rotterdam. Our results indicate a positive association between entrepreneurial intentions, BAS-D, BAS-FS, Conscientiousness, and Openness. Furthermore, we predict whether a student has 'high' entrepreneurial intention with an accuracy of 87.61% using only six personality traits. Lastly, some significant interaction effects between gender and personality were identified.

Table of Contents

1. Introduction

An individual's personality is often credited for its ability to predict or guide people towards essential life choices. For instance, when deciding on a career path, one is often encouraged to take a personality test to match one's values and personality with that of a given job (i.e., Yen, 2021). Chartrand (2016) further explains how personality is essential in the workplace, decreasing turnover and conflict, increasing worker motivation, and preventing burnouts. This concept is known more formally as the person-job fit theory, which dictates that utility increases with the match between a job and an individual's characteristics. Schneider (1987) further explains how individuals rather work for firms that have something in common with them. Therefore, a good job fit increases both job satisfaction and well-being (Kristof, 1996; Liu et al., 2010; Warr & Inceoglu, 2012).

An important, though very complex, occupation often discussed in academic literature is entrepreneurship (Gartner, 1990). Over time, many schools of thought, such as the great person and classical school, have attempted to define entrepreneurship. Definitions range from extraordinary achievers to individuals partaking in creative destruction. Cunningham and Lischeron (1991) recognize that these schools provide perceptions of the underlying values associated with entrepreneurship. They reach the consensus that the process behind entrepreneurship is relative and involves individuals bearing the risks and rewards associated with a venture. The importance of entrepreneurship in the economy becomes evident from its links to increased innovation (Zhao, 2005) and faster economic growth (Wennekers & Thurik, 1999). This importance is further reflected in the high demand for entrepreneurship in modern, globalized economies where the reallocation of resources allows for structural change. (Audretsch& Thurik, 1998; Wennekers & Thurik, 1999).

The combined importance of entrepreneurship and job fit make the relationship between personality and entrepreneurship an interesting literary field. Researchers initially failed to identify a consistent relationship and questioned whether related research had come to an empirical dead-end (Aldrich, 1999). This inconsistency led to an almost 20 year-long hiatus (Zhao et al., 2010). The hiatus further stemmed from the idea that the "ideal" psychological profile would require so many personality traits that they would contradict themselves or be a generic "Everyman" (Gartner, 1988). However, the debate on whether personality affects occupational choice re-emerged recently due to developments in meta-data analysis. Researchers opposed previous rulings, with Stewart and Roth (2007) finding a positive relationship between achievement motivation and entrepreneurship. Furthermore, various studies find an association between the Big Five Personality Traits (a.k.a. the Five Factor Model, FFM) model of personality and entrepreneurial intentions (EI) (Brandstätter, 2011; Farrukh et al., 2017; Zhao et al., 2010. Recent literature also associates EI with other models of personality, such as the Behavioural Inhibition and Behavioural Approach System (BIS/BAS) (Geenen et al., 2016; Lerner et al., 2018a). These findings suggest an association between EI and personality, disputing previous claims in the debate on personality and EI. Our paper aims to add to the debate by theorizing that if personality is associated with EI, then personality traits should at least partially predict EI. Successfully predicting EI through one's personality may thus imply an association between EI and personality. Therefore, we aim to explore both the association between EI and personality and the prediction of EI using personality traits. The paper aims to accomplish this by answering the following research question:

"Are personality traits related to entrepreneurial intentions, and can these intentions be predicted?"

The paper explores the relationship between multiple personality traits and EI using a dataset on student's entrepreneurial intentions, collected by Erasmus University Rotterdam. The selected personality traits include traits from two separate models of personality: BIS/BAS and the FFM. We select the FFM due to the significant association between the FFM and EI identified by previous academic literature (Zhao et al., 2010). Contrarily, existing research on BIS/BAS and EI is limited. However, previous literature associates various BIS/BAS sub-variables to EI (Geenen et al., 2016; Lerner et al., 2018a) and factors potentially influencing EI, such as strategic decision-making (Scheres & Sanfey, 2006). Therefore, we aim to contribute to the current literature by exploring the relationship between EI and BIS/BAS.

The paper uses multiple methodologies to answer the proposed research question. First, we use ordinary least squares regression analysis to explore the association between the personality traits of the aforementioned models and EI. Next, we use multiple classification algorithms, including Logistic Regression, Naïve Bayes, and Random Forest, to predict EI using personality traits. The prediction of EI using personality traits serves as an additional justification of the association between personality traits and EI. Lastly, previous literature implies that gender may play a moderating role in the entrepreneurial intention formation stage (Robledo et al., 2015). Therefore, we examine the moderating effect of gender on the relationship between personality and EI. This effect will be explored using moderation analysis, where the interaction effect between gender and personality traits are examined in an OLS regression with EI as the dependent variable.

The paper has multiple dimensions of relevance. Firstly, the rate of entrepreneurship for men is far larger than that of women (Acs et al., 2005). Hence, encouraging female entrepreneurship is essential to create a gender-balanced workforce. Examining the moderating effect of gender on EI allows policymakers to implement policies and programs encouraging female entrepreneurship, fostering a more balanced entrepreneurial workforce. These policy and workforce implications contribute to the research's social relevance. The paper also adds to the discussion on whether entrepreneurship is a gendered profession (Gupta et al., 2009). Furthermore, one can use the personality traits determining entrepreneurship to identify specific target groups. These groups could include students who are likely to have greater EI, who could then be encouraged or offered additional entrepreneurial education, increasing the quantity of opportunity entrepreneurship in the economy. The increased rate of opportunity entrepreneurship increases both innovation and economic growth in the economy, making the research economically relevant. Moreover, while there has been much research on the demographic determinants of entrepreneurship, knowledge on the personality determinants of entrepreneurship is limited. This lack of knowledge is emphasized by the aforementioned debate on whether personality matters for entrepreneurship (Zhao et al., 2010). The paper contributes to existing literature by using multiple methods to identify a relationship between personality traits and EI. A further contribution is the use of machine learning in economics and psychology, which remains relatively unexplored by previous academic literature. Therefore, the paper's contribution to the ongoing debate and existing literature in entrepreneurship and psychology indicates its scientific relevance.

The remaining section of this thesis is structured as followed: First, we explore the theoretical framework surrounding entrepreneurship, the relationship between EI and various personality traits, and the moderating role of gender. This theory is then used to substantiate the paper's hypotheses. Next, we discuss the origins and descriptive statistics of the dataset and the paper's dependent, independent, and control variables. This is followed by a brief explanation of the paper's two methodologies: multiple linear regression (MLR) and prediction using classification techniques. Hereafter, we discuss the results, their implications, and their limitations. Lastly, the main findings are summarized and used to answer the proposed research question.

2. Theoretical framework

This section discusses previous literature and formulates our hypotheses. First, we discuss the general relationship between personality and EI and the topic's development over time (section 2.1). Next, we discuss two personality measures: The FFM and BIS/BAS. Existing literature on both BIS/BAS (section 2.2) and the FFM (section 2.3) is reviewed and used to formulate hypotheses 1 and 2. Lastly, we discuss the moderating role of gender in section 2.4. An overview of the formulated hypotheses can be found in conceptual Models (1) and (2) in Appendix A1.

2.1. Personality and entrepreneurial intentions

Entrepreneurship involves individuals who bear the risks and rewards associated with a business venture (Cunningham & Lischeron, 1991). Miller et al. (2009) describe entrepreneurship as one of the greatest achievements of the postsecondary education system due to its positive impact on job creation and economic development. Furthermore, Zhao (2005) associated increases in entrepreneurship with greater levels of innovation, an essential factor in the economy. The large amount of research dedicated to entrepreneurship further implies its importance as a field of study. Van der Zwan et al. (2010) find that entrepreneurship develops in five stages at increasing levels of engagement. The second stage of the entrepreneurship ladder is entrepreneurial intention.

Entrepreneurial intention is an individual's conscious state of mind before carrying out an action that focuses on entrepreneurial behaviour like gaining entrepreneur status and setting up a venture (Moriano, 2012). Ajzen's (1991) theory of planned behaviour provides a conceptual framework to overcome the complexity of individual social behaviour. He finds that attitudes, perceived control, and subjective norms related to behaviour are good predictors of behavioural intentions. Together with perceived behavioural control, these intentions then account for a large part of behaviour variance, making them good predictors of behaviour. Furthermore, Kautonen et al. (2015) find that the theory of planned behaviour holds for behaviour that is related to business start-ups, such as EI. This implies that EI is likely to be a good predictor of entrepreneurship.

Previous literature reveals the importance of gender and personal attitudes (Ajzen, 1991; Davidsson, 1995), such as competitiveness and achievement in fostering EI. In addition to these personal attitudes, research in the 1980s centred around psychological characteristics (i.e., personality traits). However, the relationship between personality traits and EI was inconsistent, casting doubt on the value of trait paradigm as a determinant (Brockhaus, 1982). This doubt was reinforced by the number of traits that were found to encourage entrepreneurship. Gartner (1988) realized that an entrepreneur's ideal psychological profile requires such a large number of personality traits that they would contradict each other or make the individual a generic "Everyman."

Consequently, these outcomes led to the belief that research had reached an empirical dead-end (Aldrich, 1999). The emergence of meta-data analysis, which combines the results of multiple studies to increase the size of the dataset, reignited the field of research.

Zhao and Seibert (2006) conducted a meta-data analysis and indicated an association between the FFM and EI. Multiple papers have since found evidence suggesting the Big 5 personality traits are inputs for entrepreneurial decision-making (Brandstätter, 2011; Farrukh et al., 2017; Zhao et al., 2010). However, due to the overall inconsistency of results, the debate on whether personality affects EI is still ongoing. Recently, other personality measures such as BIS/BAS have been associated with EI, adding to the debate (Geenen et al., 2016; Lerner et al., 2018a).

2.2. BIS/BAS and entrepreneurial intentions

Gray's (1972, 1981) biopsychological theory of personality identifies two main personality dimensions: anxiety and impulsiveness. These dimensions represent the differences in sensitivity of two neurological systems between individuals. The first system, the behavioural inhibition system (BIS), is related to avoidance and withdrawal from negative and unpleasant situations (Gray 1981). It reflects the anxiety dimension along with other negative emotions. The second system is the behavioural approach system (BAS) and involves individuals moving towards a goal. Brenner et al. (2005) define BIS/BAS as crucial motivational systems for personality psychology theories and emphasize that BIS/BAS helps define an individuals' personality.

Scales have been developed to capture the two systems and thereby study differences in personality. While Gray linked BAS to positive sentiments such as happiness and hope (Gray, 1981, 1990), Carver and White (1994) realized that there is little consensus over the behavioural manifestation of BAS. Therefore, they divided the BAS scale into three sub-scales: BAS Drive, BAS Fun Seeking, and BAS Reward Responsiveness. Literature has since suggested that these sub-scales reflect constructs such as impulsivity and reward sensitivity (Caseras et al., 2003; Campbell-Sills et al., 2004; Leone & Russo, 2009).

Recently, Geenen et al. (2016) and Lerner et al. (2018a) found partial evidence implying a relationship between BIS/BAS and EI, suggesting BIS/BAS to be an input for decision-making. As BIS and BAS reflect different personality dimensions, their effects on EI differ. These relationships are further explored in sections 2.2.1 to 2.2.4.

2.2.1. BIS

The sensitivity of the BIS is related to the degree to which an individual avoids or withdraws from something unpleasant that has anxiety-related signals (Gray, 1981). Individuals with high BIS sensitivity are thus more likely to avoid situations with negative associations, such as stressful or high-risk scenarios.

Previous literature hypothesized a negative association between BIS and EI but could not support these hypotheses with statistical evidence (Geenen et al., 2016; Lerner et al., 2018a). However, Cramer et al. (2002) found that risk aversion discourages entrepreneurship choice, meaning it is probable that individuals with higher BIS scores are less likely to become entrepreneurs. Furthermore, previous literature linked BIS to anxiety, negative affect, and depressive symptoms (Campbell-Sills et al., 2004; Carver & White, 1994). Hessels et al. (2018) theorize that depression decreases the likelihood of firm survival due to reduced self-efficacy. Pihie and Bagheri (2013) further argue that self-efficacy is a strong predictor of EI for university students. Therefore, we theorize that high BIS scores and the accompanying low self-efficacy are likely to discourage EI. This leads to the first hypothesis:

Hypothesis 1a: *BIS is negatively associated with entrepreneurial intentions.*

2.2.2. BAS Drive

The second component of BIS/BAS is BAS Drive (BAS-D). Individuals with high BAS-D scores are highly motivated to follow their goals and are generally persistent in doing so. (Carver & White, 1994). Thus, these individuals are more likely to have greater levels of motivation.

Rey-Martí et al. (2015) find that motivation related to business concepts such as risk propensity (the motivation to take risks) is associated with higher entrepreneurial success rates for women. This indicates that individuals with higher levels of motivation (thus, those who are likely to have higher levels of BAS-D) are likely to have better entrepreneurial career prospects. The increased level of entrepreneurial success is likely to reflect on EI. In line with this, Lerner et al. (2018a) find that BAS-D is instrumental for the initiation of early-stage entrepreneurship. The idea is further supported by Geenen et al. (2016), who find that BAS-D has a significant positive association with EI for individuals with prior business set-up experience. Therefore, our second BIS/BAS hypothesis is:

Hypothesis 1b: *BAS Drive is positively associated with entrepreneurial intentions.*

2.2.3. BAS Fun Seeking

Individuals with greater sensitivity to BAS Fun Seeking (BAS-FS) are intrinsically motivated to find novel rewards in a spontaneous manner (Carver & White, 1994). They are more likely to seek out rewards as a spur-of-the-moment decision, implying that BAS-FS is positively related to impulsivity.

Romer et al. (2009) find that impulsivity is strongly associated with risk behaviour initiation. This implies that individuals with high BAS-FS scores may be more likely to exhibit risky behaviour, which has been linked to increased interest in entrepreneurship (Cramer et al., 2002). Furthermore, Lerner et al. (2018b) find that impulse-driven actions are required to form an efficient marketplace for entrepreneurial innovation. This is further emphasized by entrepreneur's preference for experimentation and exploration (Wang, 2008). While studies hypothesized a positive association between BAS-FS and EI, its statistical analysis failed to find a significant relationship (Geenen et al., 2016; Lerner et al., 2018a). Nevertheless, we expect a positive association between BAS-FS and entrepreneurial intention, leading to the following hypothesis:

Hypothesis 1c: *BAS Fun Seeking is positively associated with entrepreneurial intentions.*

2.2.4. BAS Rewards Responsiveness

BAS Reward Responsiveness (BAS-RR) is associated with the sensitivity an individual has towards pleasant reinforcers in their environment (Carver & White, 1994). Thus, it is an individual's reaction or anticipation towards a reward.

Franken and Muris (2005) find BAS-RR positively related to IOWA gambling test performance, which typically measures behavioural decision-making. Furthermore, previous literature reasons that financial rewards are one of the main drivers behind entrepreneurship (Kuratko et al., 1997). Therefore, individuals with high BAS-RR may have greater EI due to the potential large financial rewards related to it. However, current literature offers inconsistent findings. Geenen et al. (2016) find a negative association between BAS-RR and EI, while Lerner et al. (2018a) find a marginally positive association with EI. Lerner et al. suggest that Geenen et al.'s results stem from the high uncertainty present in the early stages of EI, meaning other activities may offer greater rewards. We theorize that the motivation to acquire large financial rewards by pursuing a career in entrepreneurship outweighs its uncertainty and reach the hypothesis:

Hypothesis 1d: *BAS Reward Responsiveness is positively associated with entrepreneurial intentions*.

2.3. The Big 5 and entrepreneurial intentions

The FFM has been the most used reference system for personality traits as of the 1980s (John et al., 2008). McDougall (1929) proposed the initial idea that personality traits can be grouped into five classes. Subsequently, systematic work by Cattel (1946, 1947) used factor-analytic studies to develop personality trait scales. These scales were later used to replicate Cattel's results, with multiple studies finding evidence for the convergence on five personality classes (Fiske, 1949; Tupes & Christal, 1961). While there was a consensus on there being five main classes, the identity of the classes was disputed. Costa and McCrae proposed (1985) the current, most commonly used FFM components: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. Previous literature indicates that The FFM's personality traits remain relatively stable for working-age individuals across time and between generations (Cobb-Clark & Schurer, 2012). Furthermore, adverse events were not found to impact personality at an economically meaningful level. These factors highlight the importance of non-cognitive skills (i.e., the FFM) as inputs for the economic decision-making process.

The quantity of research linking the FFM to factors influencing economic outcomes such as job performance (Barrick & Mount, 1991) and academic achievement (Komarraju et al., 2011) further emphasizes this importance. More recent studies also indicate a relationship between the FFM and entrepreneurial success in the form of venture survival (Ciavarella et al., 2004). Furthermore, Zhao et al. (2010) use metadata analysis to find the FFM components to be good predictors of EI. However, the nature of the relationship between each factor and EI differs. Sections 2.3.1 to 2.3.5 further explore these relationships.

2.3.1. Extraversion

Extraversion is related to how outgoing and social an individual is and is characterized by the sub traits: friendliness, gregariousness, assertiveness, activity level, excitement-seeking, and cheerfulness (Zhao et al., 2010). People with high Extraversion scores are more likely to seek out the company of others and be engaged in social activities.

Ismail et al. (2009) theorize that the extensive range of social interaction an entrepreneur faces with customers, suppliers, and employees creates a positive association with entrepreneurship for extroverts. Zhao and Seibert (2006) reinforce this, finding that entrepreneurs have higher levels of Extraversion than managers. This positive association between Extraversion and entrepreneurship may partially be explained by entrepreneurs having a more proactive personality than managers (Rauch & Frese, 2007). Furthermore, previous literature linked higher levels of Extraversion to greater interest in enterprising activities (Costa et al., 1984) and entrepreneurial intentions (Zhao et al., 2010). Therefore, the first FFM hypothesis is:

Hypothesis 2a: *Extraversion is positively associated with entrepreneurial intentions.*

2.3.2. Agreeableness

Individuals with high levels of Agreeableness are generally friendly, often put the needs of others above their own, and enjoy social harmony. These individuals are also likely to identify with the following sub-traits: trust, sympathy, altruism, cooperation, morality, and modesty (John & Srivastava, 1999).

While Agreeableness may have positive implications for teamwork, it is negatively correlated with the need for autonomy (Koestner & Losier, 1996). An example of such a need is the desire to act independently, a need satisfied by entrepreneurship. This need for autonomy implies that entrepreneurs are likely to have lower levels of Agreeableness. This is emphasized by Zhao and Seibert's (2006) findings that entrepreneurs have lower levels of Agreeableness than managers. Furthermore, Singh and DeNoble (2003) argue that entrepreneurs pursue self-interest, which opposes the altruistic view of individuals with high levels of Agreeableness. Zhao et al. (2010) further emphasize this, finding a negative association between Agreeableness and EI. Therefore, our second hypothesis is:

Hypothesis 2b*: Agreeableness is negatively associated with entrepreneurial intentions.*

2.3.3. Conscientiousness

Conscientiousness is characterized by traits including orderliness, achievement-striving, cautiousness, self-efficacy and discipline, and a sense of duty (Zhao et al., 2010). Individuals with this trait are skilled at directing their impulses to achieve a specific goal. Furthermore, high levels of Conscientiousness indicate a preference for planning and orderliness over spontaneity and are associated with greater reliability (Roberts et al., 2005).

Previous literature has found Conscientiousness to be an indicator of the urge and ability to work hard (Barrick & Mount, 1991) and job performance (Barrick et al., 2001). However, the subfactors of Conscientiousness are a source of conflict. While the achievement-striving subfactor is likely to be positively related to EI, the dutifulness sub-factor could be not or negatively related. This conflict between subfactors of Conscientiousness may lead to an overall small positive correlation between Conscientiousness and EI (Rauch & Frese, 2007). Nevertheless, multiple studies found a significant positive association between Conscientiousness and EI (Brandstätter, 2011; Zhao et al., 2010). Therefore, we form the third FMM hypothesis:

Hypothesis 2c: *Conscientiousness is positively associated with entrepreneurial intentions*

2.3.4. Neuroticism

Emotional stability implies that individuals have the ability to remain balanced, are less easy to upset, and stay calm during high-stress situations. High levels of Neuroticism indicate low emotional stability. Such individuals are often plagued with negative emotions and generally conform with the sub-traits of anger, vulnerability, self-consciousness, immoderation, anxiety, and depression (Costa & McCrae, 1992).

High levels of Neuroticism indicate vulnerability to stress (Schneider, 2004; Suls, 2001). In entrepreneurship, large amounts of pressure and stress stem from a high workload, critical decisions, and large financial consequences (Zhao et al., 2010). The vulnerability to stress associated with high Neuroticism levels may discourage such individuals from pursuing high-stress occupations like entrepreneurship. Furthermore, previous literature finds a positive association between emotional stability and EI (Zhao et al., 2010). As Neuroticism is the opposite of emotional stability, our fourth FFM hypothesis is:

Hypothesis 2d: *Neuroticism is negatively associated with entrepreneurial intentions.*

2.3.5. Openness

Lastly, Openness is characterized by the sub-traits of imagination, emotionality, artistic interests, intellect, liberalism, and adventurousness. Openness implies open-mindedness, which suggests individuals enjoy gaining new experiences and are unlikely to have a set routine. These individuals generally have higher levels of imagination and curiosity (John & Srivastava, 1999).

The greater flexibility and idea generation associated with high levels of Openness are likely to encourage new venture creation, implying the expectation of a positive relationship with EI (Ismail et al., 2009). Zhao and Seibert (2006) investigate this and find that the level of Openness of entrepreneurs is significantly higher than that of managers. Zhao et al. (2010) further enforce the significant positive association between Openness and EI. They suggest that the association can partially be explained by more creative individuals being likely to attempt a non-conventional career path. Furthermore, Openness may indicate the ability and motivation to learn (Barrick & Mount, 1991), further fostering EI. Therefore, our last FMM hypothesis is:

Hypothesis 2e: *Openness is positively associated with entrepreneurial intentions.*

2.4. Entrepreneurial intentions, personality and gender

The rate of entrepreneurship for men is far larger than that for women (Acs et al., 2005). Entrepreneurship has, therefore, long since been viewed as a gendered profession, implying gender may affect EI. Vamvaka et al. (2020) find that men are more likely to execute EI by creating new ventures. However, previous literature also finds that gender may play a moderating role in the entrepreneurial intention formation stage. Shinar et al. (2012) find that gender plays a moderating role between EI and perceived barriers to entrepreneurship, with women finding perceived entry barriers generally more important than men. The differences in perceived barriers to entrepreneurship between genders may be related to personality differences between genders. Similarly, Robledo et al. (2015) find that subjective norms, such as social pressure, are determinants of perceived behavioural control for women but not for men. This is linked to women's lower aspirations towards entrepreneurship. Due to the greater perceived barriers to entrepreneurship and the effect of social norms on EI for women, we expect that the extent to which women consider and are affected by obstacles related to entrepreneurship due to their personality traits decreases their EI. Therefore, based on these studies, it is expected that women may have a stronger negative association between personality traits discouraging EI, implying the moderating role of gender. This leads to the following hypothesis:

Hypothesis 3a: *The negative association between personality traits and entrepreneurial intentions is stronger for women than for men.*

Furthermore, Verheul et al. (2012) find that the effect of parents being entrepreneurs on the entrepreneurial status of an individual is stronger for men than for women, providing partial evidence supporting the moderating effect of gender. Similarly, Bagheri and Pihie (2014) find that entrepreneurial self-efficacy has a stronger association with EI for men than for women, again implying a stronger positive relationship between factors encouraging entrepreneurship for men than women. This leads to our last hypothesis:

Hypothesis 3b*: The positive association between personality traits and entrepreneurial intentions is less strong for women than for men.*

The positive associations include the BAS-D, BAS-FS, BAS-RR, Extraversion, Conscientiousness, and Openness personality traits. Contrarily, the negative associations include the BIS, Agreeableness, and Neuroticism personality traits. These hypotheses may explain how differences between genders in the effect of personality traits on entrepreneurial intentions contribute to entrepreneurship's gender gap.

3. Data

In this section, we first introduce our dataset (3.1). Next, in section 3.2 we explain our dependent variable, independent variables, and multiple control variables. Lastly, we look at descriptive statistics and correlations between variables in section 3.3.

3.1. The dataset

In order to explore the effect of personality and gender on entrepreneurial intentions for students, a sample of the student population at Erasmus University Rotterdam was taken. The dataset stems from a survey held by Erasmus University from May 2015 to April 2016 that included 182 university students. However, due to various missing values, the final dataset consists of 150 responses. The dataset is comparable to the dataset collected by Bernoster et al. (2018) in their study on the relationship between overconfidence, optimism, and entrepreneurship. Multiple questions in the survey have been extracted from previous literature. These statements include the questionnaires used to calculate entrepreneurial intentions (Liñán & Chen, 2009), BIS/BAS scores (Carver & White, 1994), and FFM scores (John & Srivastava, 1999), as is further discussed in section 3.2.

3.2. Variables

3.2.1. Dependent variable

The dependent variable used in our study is entrepreneurial intention, which reflects a student's intention of becoming an entrepreneur. The survey included multiple statements proposed by Liñán and Chen (2009) as part of the entrepreneurial intentions questionnaire (EIQ) to measure these intentions. The EIQ was developed using Ajzen's theory of planned behaviour and acts as a standardized measure of EI. While the complete EIQ contains 20 statements, this includes statements reflecting subjective norms, personal attitude, perceived behavioural control, and EI. Therefore, we utilize only the six statements associated with EI. Student responses to these statements were measured on a 7-point Likert scale, where 1 indicates strong disagreement and 7 strong agreement. Students' final EI scores were then calculated as the average of the six statement scores. Higher scores reflect higher EI. Empirical testing has led to the strong support of the EIQ (Liñán & Chen, 2009; Liñán & Chen, 2006), meaning that the EI calculated in our paper likely reflects actual EI. We further explore this in section 3.3 by analysing EI's Cronbach's alpha.

3.2.2. Independent variables

The independent variables consist of the components of BIS/BAS and the FFM. To measure student's scores for the four BIS/BAS components (BIS, BAS-D, BAS-FS, BAS-RR), they completed a 24 item survey with statements proposed by Carver and White (1994), translated to Dutch. BIS statements reflect the concern and sensitivity towards adverse events, while BAS statements attempt to reflect individuals' behaviour and responses towards positive or potentially rewarding situations. The extent to which the BIS/BAS scores estimated by the survey reflect student's actual BIS/BAS is explored in section 3.3 by analysing their Cronbach's alpha. Carver and White's (1994) statements are rated on a 4-point Likert scale, where 1 indicates strong agreement with the statement and 4 indicates a strong disagreement. However, in the questionnaire used for the current study, these roles are reversed, with 1 reflecting strong disagreement and 4 strong agreement. Only two statements have to be reverse scored (2 and 22), as they reflect statements opposite to higher BIS/BAS sensitivity. We calculate student's BIS score using seven statements, the BAS-RR score using five statements, and the BAS-D and BAS-FS scores using four statements. The four remaining statements are filler items and, therefore, not used. The final score of each BIS/BAS component is calculated as the average score of these statements, with higher scores reflecting a greater sensitivity to BIS/BAS traits.

The FFM sub-variable (Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness) scores are measured according to a 21 item questionnaire. Erasmus University's survey extracted these statements from the Big Five Inventory (BFI) proposed by John and Srivastava (1999). The BFI is a self-report scale that uses 44 statements to measure the FFM personality trait scores. The 21 statements selected for the survey were translated to Dutch. The survey asked students to voice their agreement with each statement on a 5-point Likert scale, with 1 being strongly disagree and 5 equalling strongly agree. Openness is measured using five statements, while the remaining FFM traits consist of four statements each. After accounting for reverse-scored questions, we calculate the final score of each FFM component as the average score of related statements. Higher FFM scores reflect a greater sensitivity towards the FFM trait in question.

3.2.3. Control variables

To reduce the omitted variable bias (OVB) in the ordinary least squares (OLS) regression, we include multiple control variables that are likely to be associated with both personality traits (BIS/BAS and FFM) and EI.

The first control is a student's age. Previous literature indicates that peak consistency in personality traits occurs after age 50 (Roberts & DelVecchio, 2000). However, even at this peak, the consistency is not sufficiently high to imply unchanging personality traits. Therefore, personality likely differs with age. Furthermore, according to Hatak et al. (2015), older employees have a lower intention to act entrepreneurially. This would imply that younger individuals have higher entrepreneurial intentions. However, individuals of middle age and older were shown to have the highest success rate (Azoulay et al., 2020). Papulová and Mokroš (2007) further emphasize this, finding that graduates often have insufficient managerial skills to become entrepreneurs. The low rates of entrepreneurial success for young individuals may thus discourage entrepreneurial intentions. Furthermore, Brockhaus (1982) finds that most entrepreneurial decisions occur between ages 25 and 40. This indicates that, while entrepreneurial intentions may increase with age initially, there may be an age after which these intentions deteriorate. This implies an inverse u-shaped relationship between age and entrepreneurial intentions. Therefore, we will control for both age and age^2. However, the extent to which the relation is u-shaped may be limited as the sample contains university students whose ages are unlikely to differ significantly.

A second important factor is whether the students' parents are entrepreneurs. Previous literature indicates that entrepreneurial parents are likely to increase a child's entrepreneurial intentions (Crant, 1996; Farrukh, 2017; Mueller, 2006). This is because entrepreneurial parents may affect how students view entrepreneurship and the extent to which entrepreneurship can lead to success. Mathews and Moser (1995) emphasize this and find that entrepreneurial parents often act as mentors for their children when becoming entrepreneurs. Furthermore, whether an individual's parents are entrepreneurs is likely to affect an individual's personality directly and indirectly. Firstly, previous literature finds that role models (i.e., entrepreneurial parents) influence individuals' beliefs and attitudes, including their perceived self-efficacy (Krueger et al., 2000). Furthermore, genetics could indirectly affect an individual's personality. Current literature implies that personality traits are partially heritable, with the average contribution of genetics to differences in an individuals' personality equalling 40% (Vukasović & Bratko, 2015). Assuming that the personality traits of parents who are entrepreneurs differ from those who are not, their children's personalities are likely to differ as well. Therefore, as an individual's parents being entrepreneurs affects their EI and personality, it should be controlled during OLS regression. Erasmus's survey indicates whether neither parent is an entrepreneur (1) and whether the father (2), mother (3) or both (4) are entrepreneurs. This categorical variable was transformed into a binary variable for our research, with either none (0) or at least one (1) of a student's parents being an entrepreneur.

Next, a student's financial status is used as a control, as capital is a common requirement for entrepreneurs. Pfeifer et al. (2016) find that family wealth is an essential factor when predicting the entrepreneurial intentions of Croatian students, with entrepreneurial intentions increasing with family wealth. Mueller (2006) similarly suggests that entrepreneurial intentions increase with gross household income. This is possibly due to low family wealth individuals preferring the financial stability of employment over entrepreneurship. Sandhu et al. (2011) further emphasize this by identifying financial resources as a significant barrier to entrepreneurial intentions for students. Furthermore, previous literature identifies a positive association between household income, the development of a child's positive personality traits, and behavioural and emotional health (Akee et al., 2018). Therefore, the personalities of students with different financial statuses are likely to differ. As an individual's financial status is related to both EI and personality, it should be controlled for in OLS regression. Although the wealth of student's families was not recorded in the survey, we will use whether the student has a job to indicate family wealth. Those with a poorer family background are less likely to be financially supported throughout university, making them more likely to have a job besides their study.

Lastly, previous literature has repeatedly implied an association between gender, personality traits, and EI (Liñán et al., 2011; Minniti & Nardone, 2007), emphasizing its importance as a control variable. Furthermore, as explained in section 2.4, gender is likely to have a moderating effect on entrepreneurial intentions. Therefore, gender will be used both as a control and for moderation purposes. For this variable, one indicates female, while zero indicates a male student.

3.3. Descriptive Statistics

The data's descriptive statistics show that respondents' gender is relatively balanced, with 55% of respondents being female (Table 1). Students participating in the questionnaire had an average age of 20.6 years. Furthermore, 54.7% of students reported having a job, while 29.3% reported that at least one of their parents is an entrepreneur. On average, students' entrepreneurial intentions are 3.28, while the sum of FFM scores is 17.12. Moreover, the average sum of BIS/BAS is 12.05, which is comparable to previous studies' average sum of BIS/BAS score (e.g., Geenen et al., 2016; Lerner et al., 2018a). Lastly, we use Cronbach's alpha to assess the internal consistency and reliability of the variables derived from combining Likert scale questions. EI has an excellent internal consistency (0.94), while most BIS/BAS and FFM components show either good or acceptable consistencies (ranging from 0.685 to 0.819). BAS-FS (0.574) and BAS-RR (0.541) both have poor internal consistencies, while Agreeableness (0.488) is below the acceptable threshold (0.50). Therefore, it is unlikely to reflect Agreeableness accurately. The low scores could be caused by the small number of questions per variable (4) or a lack of correlation between questions.

Table 2 presents the Pearson correlations between variables, which indicate the magnitude and direction of the association. Previous literature implied that the three components of BAS are collinear as they all reflect behavioral activation (Willem et al., 2010). Similarly, we find the three measures of BAS to be collinear, as well as BAS-RR and BIS (0.26). These correlations decrease the precision of our coefficient estimates as well as the model's statistical power. However, as none of the correlations exceeded 0.80, they are not severe enough to invalidate our results.

Furthermore, the FFM components do not appear to be collinear, except for the minor correlation between Extraversion and Openness (0.20). They are, however, significantly correlated with the BIS/BAS components. The BAS components of BIS/BAS appear to be correlated with Extraversion, Conscientiousness, and Openness. However, their correlations do not exceed 0.80. Contrarily, the correlation between BIS and Neuroticism is 0.78. This severe collinearity between explanatory variables likely affects the outcome of OLS. Therefore, when testing hypotheses 1 and 2, both models (BIS/BAS and FFM) are estimated separately first to explore their relationship with EI.

Furthermore, being female is significantly positively associated with BAS-RR, BIS, Extraversion, Neuroticism, and Openness scores. Lastly, EI appears to be positively correlated with all BAS measures, Extraversion, Conscientiousness, Openness, age, and having entrepreneurial parents, while being negatively associated with BIS and Neuroticism. These relationships are further investigated using regression analysis and classification techniques in section 4.

Table 1*: Descriptive statistics of students at Erasmus University Rotterdam in the dataset*

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**Cronbach's alpha**
BAS-D	2.973	0.487	1.5	4	0.703
BAS-FS	2.787	0.526	2.75	4	0.574
BAS-RR	3.396	0.361	2.4	4	0.541
BIS	2.896	0.564	1	4	0.819
Extraversion	3.580	0.708	1.75	5	0.764
Agreeableness	3.370	0.614	1.5	4.75	0.488
Conscientiousness	3.632	0.647	1.25	5	0.685
Neuroticism	2.810	0.872	1	5	0.813
Openness	3.725	0.735	1.8	5	0.733
Age	20.640	2.064	18	30	
Gender	0.55	0.500	0	1 = female	
Job	0.547	0.499	0	1 = Has a job	
Ent. Parents	0.293	0.457	0	1 = parents are entrepreneurs	
Ent. Intentions	3.282	1.592	1	7	0.949

Table 2*: Correlation matrix for all variables*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) BAS-D	1.00												
(2) BAS-FS	0.32***	1.00											
(3) BAS-RR	0.52***	0.36***	1.00										
(4) BIS	-0.11	-0.074	0.26***	1.00									
(5) Extraversion	0.29***	0.31***	0.36***	-0.02	1.00								
(6) Agreeableness	-0.09	-0.10	0.02	0.03	0.06	1.00							
(7) Conscientiousness	0.38***	-0.14*	0.25***	-0.04	0.12	0.08	1.00						
(8) Neuroticism	-0.10	-0.08	0.16**	0.78***	-0.04	-0.12	-0.07	1.00					
(9) Openness	0.22***	0.45***	0.26***	0.03	0.20**	-0.03	-0.08	0.01	1.00				
(10) Age	0.09	0.15*	0.07	-0.02	0.10	-0.10	-0.07	-0.07	0.04	1.00			
(11) Gender	0.05	0.02	0.17**	0.37***	0.19**	0.08	0.04	0.36***	0.22***	-0.01	1.00		
(12) Job	0.21***	0.08	0.04	-0.11	0.21**	-0.21**	0.06	-0.03	0.12	0.08	0.15*	1.00	
(13) Ent. Parents	0.17**	-0.07	0.10	0.06	0.01	-0.10	0.18**	0.04	0.09	0.08	0.17**	0.12	1.00
(14) Ent. Intentions	0.37***	0.34***	0.24***	-0.14*	0.15*	-0.12	0.17**	-0.16**	0.28***	0.16**	-0.02	0.11	0.23***

Note. Columns (1) to (13) show the Pearson correlation between variables; * = p<0.1, ** = p<0.05, *** = p<0.01.

4. Methodology

4.1. Multiple linear regression

Ordinary Least Squares (OLS) regression analysis was chosen as the method of analysis as EI is treated as a continuous variable. The regression analysis works by minimizing the sum of squared differences between a variable's predicted and actual values. We use STATA MP 15.0 to carry out the regression analysis. OLS has multiple assumptions that may severely impact results when violated. The first assumption is that of random sampling. The nature of the research data is slightly problematic, as surveys can suffer from self-selection bias, with students participating in the survey differing from those who do not. Secondly, homoscedasticity assumes that the variance of error terms is consistent. Violating this assumption leads to biased standard errors (SEs), and hence, to inaccurate hypothesis tests and significance levels. A Breusch-Pagan test was conducted and indicated that most regression models suffer from heteroskedasticity (See Appendix A7). Therefore, we use robust SEs in regressions.

Furthermore, the OLS regression variables should not be collinear. As explained in section 3.3, the collinearity between BIS and Neuroticism (0.78) violates this assumption. Therefore, hypotheses 1 and 2 are tested in separate models. After separation, a VIF analysis (See Appendix A8) did not identify further severe multicollinearity in the models. Thus, we assume that the assumption is no longer violated. Lastly, OLS's exogeneity assumption may be violated due to the exclusion of control variables affecting both the dependent and independent variables. As it is impossible to include all control variables in the model, this assumption is likely to be violated. This leads to omitted variable bias (OVB) and creates a biased EI coefficient, violating the conditional independence assumption. Therefore, as the coefficients cannot be interpreted as causal relations, we will interpret them as associations instead.

To determine which personality traits are significantly associated with EI, we conduct an OLS regression analysis with EI as the dependent variable and personality traits (BIS/BAS and FFM) as independent variables. We include multiple control variables to limit the model's OVB. As multiple variables are measured on Likert scales of different sizes (BIS/BAS components on a 1 to 4 scale, FFM components on a 1 to 5 scale, and EI on a 1 to 7 scale), the OLS coefficients are transformed to standardized coefficients for comparison purposes and are displayed in a separate column. If the p-value of an OLS coefficient is below 5%, we consider the trait to be significantly associated with EI. This will aid in answering hypotheses 1 and 2. The mathematical models for H1 and H2 are shown below. β_0 is the constant term while β_1 to β_{10} are the coefficients of the independent variables.

(H1) $$EI = \beta_0 + \beta_1 BAS_D + \beta_2 BAS_{FS} + \beta_3 BAS_{RR} + \beta_4 BIS + \beta_5 Female + \beta_6 Age + \beta_7 Age^2 + \beta_8 Ent.Parents + \beta_9 Job$$

(H2) $EI = \beta_0 + \beta_1 Extraversion + \beta_2 Agreeableness + \beta_3 Conscientiousness + \beta_4 Neuroticism + \beta_5 Openness + \beta_6 Female + \beta_7 Age + \beta_8 Age^2 + \beta_9 Ent.Parents + \beta_{10} Job$

To determine whether gender moderates the effect of personality on entrepreneurial intentions, we repeat the previous OLS process while adding interaction effects between gender and personality traits. The standardised coefficients are again added as an additional column. To answer hypotheses 3a and 3b, we analyse whether the interaction effects are statistically significant. Two models are constructed, one for the interaction effect between gender and BIS/BAS and the other for gender and the FFM. The mathematical models for these effects are displayed below. X_1 to X_n imply the personality traits, while * indicates interaction effects between variables.

(H3a/b) $EI = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \beta_{n+1} Female * X_1 + \beta_{n+2} Female * X_2 + \cdots \beta_k Female * X_n + \beta_{k+1} Female + \beta_{k+2} Age + \beta_{k+3} Age^2 + \beta_{k+4} Ent.Parents + \beta_{k+5} Job$

4.2. Classification techniques

This section explores the methodology for the different classification techniques used to predict entrepreneurial intentions. Multiple classification techniques are used as the best performing classifier may depend on the dataset and the nature of the variables. The program used for classification is WEKA, version 3.8.4. Previous Literature used Naïve Bayes to predict FFM personality traits using an individual's Twitter texts (Pratama & Sarno, 2015). Furthermore, multiple classifiers have been used to identify entrepreneurial individuals using demographic variables (Montebruno et al., 2020). However, no previous research used personality traits to predict EI using classification.

The continuous EI variable is rounded to the nearest integer, transforming it into a categorical variable and enabling classification. As EI is measured on a Likert Scale, it is an ordinal categorical variable. However, multinomial classification techniques do not utilize a variable's order information, leading to a loss of data. Therefore, we convert the proposed classification models to ordinal classification using WEKA's ordinal class classifier package, as proposed by Frank and Hall (2001). Alternatively, we also convert EI into a binary variable. Even though a data loss occurs, the accuracy of prediction for binary classification is significantly higher. While this is partially due to a higher probability of correct random classification, it is also caused by the relatively bigger sample size and hence, the training set size of each outcome. To convert EI to a binary variable, (0) will be considered low or no EI, while (1) indicates high EI. High EI is composed of individuals with EI scores from 5 to 7. EI scores of 4 are included in low EI as they are neutral responses, thus do not imply high EI. The two methods (binary and ordinal) are discussed and compared in section 5.1.

4.2.1. Imbalanced dataset

When converting EI to a binary variable, a severe data imbalance occurs: 117 of the students in the sample showed low entrepreneurial intentions while only 33 exhibited high EI. This likely severely decreases the validity and credibility of the classification algorithms, as an accuracy of 78% (117/150) can be achieved simply by classifying all students as “low EI”. Therefore, a Majority Weighted Minority Oversampling Technique (MWMOTE) generates 84 new instances of students with high EI. MWMOTE is a modification of the synthetic minority oversampling technique that is less sensitive to noisy instances proposed by Barua et al. (2012). It identifies informative minority class (low EI) samples and allocates weights relative to the sample’s Euclidean distance to the closest majority class (high EI) sample. This occurs until no data points can be found outside these clusters. The weights indicate the number of new instances that a cluster generates. The newly generated instances are then combined with the original data to create a more balanced dataset consisting of 234 samples. The R-code for the implementation of MWMOTE is presented in Appendix A6.

4.2.2. Performance measures

The performance measures used include accuracy and F1 score. A classifier’s accuracy can be described as its ratio of correctly versus incorrectly classified instances. However, accuracy’s sensitivity to skewed data may be problematic if our independent variables (personality traits) are severely skewed. Therefore, we also measure the F1 score, a more robust measure for skewed data. A classification technique’s F1 score is the weighted average value of its precision (ratio of people correctly classified as high EI to the total number of individuals predicted to have high EI) and recall (ratio of correctly predicted high EI individuals to the actual number of people with high EI).

4.2.3. Attribute Selection

Each classification method may allocate different levels of importance to the BIS/BAS and FFM variables, making them either more or less applicable to the model. The WrapperSubsetEval method proposed by Kohavi and John (1997) is used to select relevant personality traits for the prediction of EI for each classifier. The method creates a cross-validation loop by wrapping a classier and searching for the combination of attributes that result in the highest model accuracy (as estimated using accuracy estimation techniques). Furthermore, we use a greedy best-first search mechanism with a forward search direction as it increases the accuracy of classifiers (Kohavi, 1994). Furthermore, it overcomes small local minima formed by feature interaction, such as correlations between personality traits (Kohavi & Sommerfield, 1995). After the attribute selection process, we remove any variables not selected by the wrapper.

4.2.4. Logistic Regression

The first classification technique, Logistic regression (LR), is a predictive analysis algorithm with the sigmoid function (an S-shaped function with values between 0 and 1) as its core. The method weighs its input variables (personality traits) according to the value of their coefficients, producing a binary output value based on probability. WEKA runs the classifier and predicts in which class (high or low EI) a student is placed. The classification algorithm was chosen as it works well with a binary dependent variable, making it a potential predictor of EI (binary version), where students have either high (1) or low (0) entrepreneurial intentions. Similar to OLS, LR comes with multiple assumptions: there should not be any severe multicollinearity between predictors or any severe outliers. As mentioned in section 3.3, there may be some collinearity issues. However, attribute selection is likely to help fulfil the assumption by removing collinear predictors.

4.2.5. Naïve Bayes

The second classifier is Naive Bayes, which is an algorithm developed from Bayes Probability Theorem. Naïve Bayes splits the dataset into two parts: a feature matrix composed of vectors of dependent variables values (personality traits scores) and a response vector with the predicted outcome (high or low EI) of every vector (student). Naïve Bayes assumes that variables are independent and carry equal weights towards prediction. Bayes' probability theorem is fused with the independence assumption and the maximum a posteri rule to create the Naïve Bayes classifier. Finally, WEKA is used to run the classifier. Naïve Bayes was chosen as it works well with nominal and binary dependent variables. Furthermore, the classifier does not require a large model size or training set due to its fast convergence, making it applicable to our relatively small dataset of 150 students. A further advantage of the classifier is that it cannot model complex behaviour. This greatly reduces the probability of overfitting, which is a crucial problem in classification (Dietterich, 1995).

4.2.6. Random Forest

The Random Forest (RF) classification method randomly selects a subset of the training dataset and uses it to generate various Decision Trees. The Decision Trees each predict the outcome of classification and are combined to create a stable prediction using a bagging algorithm. Thus, the outcome with the most votes becomes the RF's predicted outcome. Unlike the other two classifiers, RF is robust to noise, which is likely to occur due to the nature of self-surveys. Furthermore, the algorithm is characterized by low levels of overfitting and variance.

5. Results

5.1. Regression analyses

The following section aims to find statistical evidence for Hypotheses 1 to 4 using the OLS regression analysis provided in Tables 3a and 3b. Model 1 in Table 3a presents a model only including the control variables. Model 2 tests BIS/BAS-related hypotheses 1a to 1d, while Model 3 tests FFM hypotheses 2a to 2e. Model 5 analyses the moderating effect of gender (H3a and H3b) on BIS/BAS and EI, while Model 6 tests the effect of gender on the FFM and EI. Appendix A2 provides an overview of the associations found in the regression analysis.

5.1.1. Hypothesis 1: BIS/BAS

Hypothesis 1a to 1d hypothesise the relationship between the four BIS/BAS components and entrepreneurial intentions. These hypotheses are tested in Model 2 in Table 3a. The first hypothesis (H1a) theorized that BIS is negatively associated with entrepreneurial intentions. While Model 2 indicates a negative BIS coefficient, the coefficient is not significant at the 5% level. Thus, as we cannot confirm that a higher BIS score is associated with lower EI, H1a is rejected. The second hypothesis (H1b) argues that BAS-D is positively associated with EI. Model 2 shows that BAS-D has a positive standardized coefficient of 0.224 that is statistically significant at the 5% level. Thus, an increase of BAS-D by one standard deviation (SD) leads to an average increase of EI by 0.224 SDs, ceteris paribus. As the coefficient is statistically significant, H1b cannot be rejected.

Furthermore, Hypothesis 1c (H1c) speculated that BAS-FS is positively associated with EI. Model 2 indicates a positive BAS-FS coefficient (β = 0.264) that is statistically significant at the 1% level. Thus, a one SD increase in BAS-FS increases entrepreneurial intentions by 0.264 SDs, ceteris paribus. Therefore, H1c cannot be rejected. Lastly, Hypothesis 1d postulated a positive association between BAS-RR and EI. However, while the coefficient was positive, it was not statistically significant at the 5% level. This means that H1d is rejected.

5.1.2. Hypothesis 2: FFM

Hypothesis 2a to 2e were created to explore the relationship between the FFM components and EI and were tested in Model 3. Hypothesis 2a theorized a positive relationship between Extraversion and EI. While the coefficient of Extraversion was positive, it was not significantly at the 5% level. Thus, H2a is rejected. Similarly, while H2b hypothesized a negative association between Agreeableness and EI, the negative association identified in our regression is statistically insignificant. Therefore, H2b is also rejected.

The third FFM hypothesis (H2c) implied a positive association between Conscientiousness and EI. Model 3 shows a positive association (β = 0.163) that is statistically significant at the 5% level. An increase in Conscientiousness by one SD thus increases EI by 0.163 SDs, ceteris paribus. As the association is statistically significant, H2c cannot be rejected. Hypothesis 2d stipulated a negative relationship between Neuroticism and entrepreneurial intentions. However, the negative association found in our regression was not statistically significant. Hence, H2d is rejected due to a lack of a significant association. Lastly, hypothesis 2e (H2e) postulated a positive relationship between Openness and EI. Model 3 indicates a coefficient of 0.252, which is significant at the 1% level. This means that an increase in Openness by one SD increases E1 by 0.252 SDs, ceteris paribus. Therefore, H2e cannot be rejected.

5.1.3. Hypothesis 3: Moderation analyses

Hypotheses 3a and 3b stipulated a moderating effect of gender on the relationship between personality traits and EI. Hypothesis 3a predicted that the negative association between personality traits and EI is stronger for women. This means that the interaction effect between gender and personality traits negatively associated with EI is negative. However, neither the interaction effect between gender and BIS, Agreeableness, nor Neuroticism were statistically significant. Therefore, hypothesis 3a is rejected. Hypothesis 3b stipulated that the positive association between personality traits and EI is less strong for women. Thus, the interaction effect between gender and personality traits positively related to IE should be negative. While the interaction effects between gender and BAS-FS, BAS-RR, and Extraversion were negative, they were not statistically significant. Furthermore, the negative interaction effect between gender and Openness (β = -0.776) was significant at the 10% level yet not at the 5% level and thus cannot be considered statistically significant. Lastly, contrary to H3b, the interaction effect between gender and BAS-D was positive (β = 1.081) and significant at the 5% level. Thus, for women, the increase in EI associated with an increase in BAS-D of one SD is 1.081 SDs larger. With this being the only statistically significant interaction effect, we do not consider gender to have a moderating role between personality traits and entrepreneurial intentions for students in the sample.

5.1.4. Controls

Neither gender nor whether a student has a job were found to affect EI significantly. Contrarily, Age is found to be significantly negatively associated with EI, while Age^2 has a positive association, implying the aforementioned inversed parabolic relationship. Furthermore, similar to previous literature, having entrepreneurial parents significantly increases a student's EI.

Table 3a: *OLS regression result with EI as the dependent variable*

Independent variable	(Model 1)			(Model 2)			(Model 3)			(Model 4)		
	B	**β**	**SE**	**B**	**β**	**SE**	**B**	**β**	**SE**	**B**	**β**	**SE**
BIS/BAS												
BAS-D				0.734**	0.224**	0.327				0.565*	0.173*	0.340
BAS-FS				0.799***	0.264***	0.250				0.764***	0.253***	0.287
BAS-RR				0.119	0.027	0.412				0.002	0.000	0.421
BIS				-0.274	-0.097	0.276				0.045	0.016	0.408
FFM												
Extraversion							0.176	0.079	0.179	-0.038	-0.017	0.188
Agreeableness							-0.267	-0.108	0.196	-0.170	-0.069	0.193
Conscientiousness							0.402**	0.163**	0.211	0.310	0.126	0.235
Neuroticism							-0.271	-0.148	0.175	-0.250	-0.137	0.237
Openness							0.546***	0.252***	0.177	0.237	0.110	0.203
Control variables												
Female	-0.196	-0.061	0.260	-0.105	-0.033	0.263	-0.187	-0.059	0.291	-0.107	-0.034	0.286
Age	-1.463**	-1.897**	0.607	-1.689***	-2.190***	0.593	-1.306**	1.693**	0.584	-1.628***	-2.110***	0.578
Age^2	0.036***	2.040***	0.014	0.041***	2.277***	0.014	0.032**	1.817**	0.013	0.039***	2.201***	0.013
Ent. Parents	0.816***	0.234***	0.279	0.788***	0.226***	0.254	0.653**	0.187**	0.278	0.692**	0.199**	0.268
Job	0.226	0.071	0.258	-0.035	-0.011	0.250	-0.018	-0.006	0.241	-0.058	-0.018	0.249
Constant	17.574***		6.684	16.500**		6.788	13.745**		6.909	15.340**		7.045
Observations	150	150	150	150	150	150	150	150	150	150	150	150
R^2	0.112			0.287			0.233			0.315		
Adjusted R^2	0.081			0.242			0.178			0.244		

Note. B= Unstandardized coefficient, β = standardized coefficient, SE= Standard error; * = p<0.1, ** = p<0.05, *** = p<0.01.

Table 3b: *OLS regression result with EI as the dependent variable*

Independent variable	(Model 5)			(Model 6)			(Model 7)		
	B	**β**	**SE**	**B**	**β**	**SE**	**B**	**β**	**SE**
BIS/BAS									
BAS-D	0.150	0.046	0.379				0.054	0.016	0.409
BAS-FS	1.159***	0.383***	0.326				0.819*	0.271*	0.418
BAS-RR	0.349	0.079	0.473				0.073	0.016	0.513
BIS	-0.727**	-0.258**	0.350				-0.349	-0.124	0.463
FFM									
Extraversion				0.285	0.127	0.212	0.173	0.077	0.227
Agreeableness				-0.105	-0.042	0.229	0.047	0.019	0.249
Conscientiousness				0.131	0.053	0.258	0.123	0.050	0.308
Neuroticism				-0.355	-0.194	0.230	-0.208	-0.113	0.291
Openness				0.823***	0.394***	0.212	0.559**	0.259**	0.253
BIS/BAS interaction									
Female*BAS-D	1.119**	1.081**	0.612				1.058*	1.022*	0.760
Female*BAS-FS	-0.687	-0.626	0.461				-0.155	-0.141	0.574
Female*BAS-RR	-0.537	-0.586	0.811				-0.311	-0.339	0.630
Female*BIS	0.871	0.868	0.544				0.774	0.771	0.899
FFM interaction									
Female*Extraversion				-0.181	-0.217	0.359	-0.270	-0.323	0.368
Female*Agreeableness				0.201	-0.224	0.379	-0.263	-0.293	0.385
Female*Conscientiousness				0.381	0.453	0.384	0.168	0.200	0.429
Female*Neuroticism				0.132	0.138	0.326	-0.066	-0.069	0.435
Female*Openness				-0.620*	-0.776*	0.350	-0.703	-0.881	0.428
Control variables									
Female	-2.176	-0.682	2.554	1.684	0.528	2.814	0.025	0.008	2.922
Age	-1.566***	-2.030***	0.562	-1.215**	-1.575**	0.569	-1.500***	-1.944***	0.552
Age^2	0.038***	2.146***	0.013	0.030**	1.700**	0.013	0.0366***	2.050***	0.012
Ent. Parents	0.717***	0.206***	0.250	0.614**	0.176**	0.278	0.633**	0.182**	0.270
Job	-0.018	-0.006	0.250	0.052	0.016	0.256	0.022	0.007	0.263
Constant	16.125**		6.683	11.937*		6.779	13.979**		6.823
Observations	150	150	150	150	150	150	150	150	150
R^2	0.326			0.264			0.372		
Adjusted R^2	0.262			0.182			0.257		

Note. B= Unstandardized coefficient, β = standardized coefficient, SE= Standard error; * = p<0.1, ** = p<0.05, *** = p<0.01.

5.1.5. Model overview

The model with the lowest adjusted R^2 (0.081) is Model 1, which only includes control variables. This emphasizes the importance of the effect of personality on EI, as models including personality traits have larger R^2 values, even when controlling for the number of variables included. The model with the highest adjusted R^2 (0.262) is Model 5, which includes BIS/BAS and the interaction effects between gender and personality traits. Furthermore, Model 2 has a larger adjusted R^2 than Model 3, indicating that BIS/BAS can explain more of the variation in EI than the FFM. Model 4 incorporates both BIS/BAS and FFM. However, the increase in adjusted R^2 from Model 2 to Model 4 is only 0.02, indicating that adding the FFM does not significantly increase the explanatory power of the regression.

5.2. Predicting entrepreneurial intentions

In this section, the results of the classification techniques used to predict EI are discussed. Three datasets are used to predict EI. The first is a 'Binary EI' dataset where EI is either low or high. Next is a 'Balanced Binary EI' dataset where additional instances of high EI are added to the dataset using MWMOTE (see section 4.2.1). Last is the 'Ordinal EI' dataset, where EI is a categorical variable (1 to 7). For each dataset, we apply three classification methods (NB, LR, and RF). Table 5 indicates the accuracies and F1 scores of each classification method. We then use the WrapperSubsetEval method (see section 4.2.3) to select attributes for each dataset and classification method, which are displayed in Table 4. After deleting variables that were not selected, the performance measures of each dataset and classification method are re-estimated, as indicated by an Asterix in Table 5.

While Table 5 indicates that the accuracies of classifiers for the unbalanced binary EI dataset are relatively large at 82.00% for LR, 82.33% for NB, and 76.67% for RF, they will not be considered significant due to the data imbalance. As mentioned in section 4.2.1, the classifiers can achieve a 78% (117/150) accuracy by classifying all students as "low EI". Therefore, the classifiers only increase the predictive accuracy of classification by 4.00%, 3.33%, and 0.00%, respectively. Contrarily, the 'Balanced Binary EI' dataset classification shows significant classifier accuracies and F1 scores. Initially, the predictive accuracies of Logistic Regression (75.21%) and Naïve Bayes (79.06%) appear to decrease relative to the unbalanced dataset. However, after accounting for the accuracy achieved by allocating all students to 'low EI' (78% for the unbalanced dataset and 50% (117/234) for the balanced dataset), their accuracies have actually increased by 21.21% and 25.73%, respectively. Furthermore, RF's accuracy has increased significantly to 85.47%. This further increases to 87.61% when including only selected attributes. Table 4 (column 2, row 3) indicates that these attributes include Extraversion, Openness, and all BIS/BAS variables. Hence, given a student's Extraversion, Openness, and BIS/BAS scores, we can correctly predict whether they have high or low entrepreneurial intentions in 87.61% of instances. The validity of RF's 87.61% accuracy is supported by its corresponding F1 score (0.876).

Next, we estimated the accuracies of the ordinal categorical dataset. These were, as excepted, significantly lower than the binary dataset accuracies. This is because a 50% chance of correctly classifying instances at random arises in binary classification, compared to a 14.29% chance when having seven categories (1/7 ≅ 14.29%). These random chance components have been subtracted from the classification accuracies in columns labelled (2) in Table 5. However, even when disregarding chance components, the accuracies of prediction of the ordinal categorical dataset were significantly lower than those of the balanced binary dataset. The largest accuracy is that of Logistic regression with only relevant attributes selected (32.00%). These attributes include Extraversion, Agreeableness, and BAS-D. Thus, given student's scores for these traits, we can correctly identify their level of EI on a scale of 1 to 7 in 32.00% of the instances. When comparing the best functioning classifier of the Balanced Binary and the ordinal categorical dataset, the accuracy of classification of the Balanced Binary dataset (87.61%) was significantly larger than that of the Ordinal categorical dataset (32.00%). This holds even after accounting for random chance components, where their accuracies are reduced to 37.61% and 17.71%, respectively. Thus, we recommend the Binary Balanced Dataset for further analysis.

Section 5.1.5 identified that a model incorporating only control variables has a lower explanatory power than a model that included personality traits. Therefore, we classified EI using the control variables proposed in section 3.2.3 (See Appendix A3). The classification method with the largest predictive accuracy was the RF classifier, with an accuracy of 56.41%. This is significantly lower than the 86.61% accuracy achieved using personality traits, further implying the importance of personality in the prediction of EI.

Furthermore, in appendices A4 and A5, classification accuracies are presented for EI classification using the BIS/BAS and FFM models separately. The largest predictive accuracy for BIS/BAS was achieved by an RF classifier with an accuracy of 83.33%, while the FFM achieved an 80.34% accuracy, also using an RF classifier. Notably, these accuracies are surpassed when using attributes from both models (87.61%), as discussed before. Similar to our findings in section 5.1.5, the BIS/BAS model appears to have a higher predictive accuracy (and explanatory value) than the FFM model.

Table 4: *Attribute selection for each classifier and dataset*

Classifier	Binary	Balanced Binary	Ordinal
Logistic Regression	Extraversion	Extraversion	Extraversion
	Neuroticism	Conscientiousness	Agreeableness
	Openness	Openness	BAS-D
	BAS-D	BAS-D	
		BAS-RR	
Naïve Bayes	Agreeableness	Conscientiousness	Extraversion
	Neuroticism	Neuroticism	Neuroticism
	Openness	Openness	BAS-FS
	BAS-D	BAS-D	
	BAS-FS	BAS-FS	
		BAS-RR	
		BIS	
Random forest	Neuroticism	Extraversion	Extraversion
	Openness	Openness	BAS-FS
	BAS-FS	BAS-D	
	BAS-RR	BAS-FS	
	BIS	BAS-RR	
		BIS	

Note. The columns indicate the dataset used for the selection of attributes; For the ordinal dataset, attribute selection took place with the ordinal class classifier version of each of the classification techniques.

Table 5: *Classification results for the BIS/BAS and FFM variables with respect to entrepreneurial intentions (EI)*

Classifier		Logistic Regression		Naïve Bayes		Random forest	
		(1)	(2)	(1)	(2)	(1)	(2)
Binary EI	Accuracy	82.00%	4.00%	81.33%	3.33%	76.67%	-1.33%
	F1	0.806		0.813		0.734	
Binary EI*	Accuracy	83.33%	5.33%	84.00%	6.00%	80.00%	2.00%
	F1	0.817		0.826		0.779	
Balanced Binary EI	Accuracy	75.21%	25.21%	79.06%	29.06%	85.47%	35.47%
	F1	0.752		0.790		0.855	
Balanced Binary EI*	Accuracy	77.35%	27.35%	79.06%	29.06%	87.61%	37.61%
	F1	0.773		0.791		0.876	
Ordinal EI	Accuracy	24.67%	10.38%	27.33%	13.04%	23.33%	9.04%
	F1	0.224		0.247		0.220	
Ordinal EI*	Accuracy	32.00%	17.71%	28.67%	14.38%	28.67%	14.38%
	F1	NA**		0.287		0.285	

Note. (1) includes the model's accuracy, while (2) subtracts the classification accuracy achieved by classifying all students as 'low EI,' which is 78% for the 'Binary EI' dataset, 50% for the 'Balanced Binary EI' dataset and an average of 14.29% for the 'Ordinal EI' dataset (scale = 1-7, 1/7 ≅ 14.29%); * indicates that only attributes selected by the WrapperSubsetEval method have been included during classification; For the ordinal dataset, the attribute selection took place with the ordinal class classifier version of each of the classification techniques; NA** indicates that one or more of the components of the F1 score cannot be calculated, meaning the F1 score cannot be calculated and is not displayed.

6. Discussion and conclusion

6.1. Findings

This paper aimed to explore the association between different personality traits and EI. We use OLS regression analysis to explain EI and multiple classification methods to predict EI in association with two models of personality: BIS/BAS and the FFM. We find statistical evidence for the association between personality and EI and the ability of personality traits to predict EI accurately. Hereby, we contribute to the debate on the importance of personality in determining EI.

First, we explored the relation between EI and BIS/BAS using regression analysis. Similar to Geenen et al. (2016) and Lerner et al. (2018a), we found a positive association between BAS-D and EI. Thus, individuals that are more motivated and driven to achieve their goals are likely to have greater EI. This is possibly due to the positive association between business-concept-related motivation and entrepreneurial success, which likely increases EI (Rey-Martí et al., 2015). Furthermore, we identify a positive association between BAS-FS and EI, meaning that individuals that spontaneously seek new rewards are likely to have greater EI. While previous studies did not identify this association (Geenen et al., 2016; Lerner et al., 2018a), it likely stems from the positive association between BAS-FS scores and impulsivity. Previous literature has associated impulsivity with risk-taking behaviour and an increased interest in entrepreneurship (Cramer et al., 2002; Romer et al., 2009). Unlike previous literature, our regression analysis did not identify an association between EI and BAS-RR. Albeit insignificant, the association was positive, similar to findings by Lerner et al. (2018a). This disputes the negative association identified by Geenen et al. (2016). Furthermore, similar to previous literature (Geenen et al., 2016; Lerner et al., 2018a), BIS was not significantly associated with EI.

Furthermore, the FFM regression analysis results indicate that only Conscientiousness and Openness are significantly associated with EI. Similar to findings by Brandstätter (2011) and Zhao et al. (2010), we find that Conscientiousness is positively associated with EI, indicating that individuals prone to planning and directing themselves to a goal are likely to have greater EI. This disputes Rauch and Frese's (2007) concerns over a conflicting association between Conscientiousness and EI. Furthermore, the association between Openness and EI was significantly positive, complementing findings by Zhao et al. (2010). Open-minded, creative individuals that enjoy engaging in new activities are thus likely to have greater EI. Unlike results from previous literature (Costa et al., 1984; Zhao et al., 2010; Zhao & Seibert, 2006), Extraversion, Agreeableness, and Neuroticism are not significantly associated with EI. The lack of significant results possibly stems from our study's limited 150-student sample size compared to previous studies' use of meta-data analysis.

The moderation analysis conducted garnered insufficient evidence for a moderating effect of gender on the relation between EI and personality traits. Unlike previous findings of gender moderating perceived barriers to entrepreneurship (Shinar et al., 2012), we did not find evidence for the negative association between personality traits and EI being stronger for women. Furthermore, while previous literature identified that the positive association between EI, entrepreneurial parents, and self-efficacy is less strong for women (Bagheri & Pihie, 2014; Verheul et al., 2012), our analysis could not produce similar results concerning the positive association between personality traits and EI. On the contrary, the interaction effect between gender and BAS-D was significantly positive (1.08). A greater BAS-D score was thus significantly associated with EI for women but not for men. Previous literature identified that entrepreneurship is often viewed as a gendered, masculine profession (Gupta et al., 2009). Furthermore, Robledo et al. (2015) find that subjective norms are a determinant of perceived behavioural control and EI for women. Thus, the subjective norms associated with gender stereotypes may discourage women from developing EI. However, increased motivation and drive towards one's goals may negate the extent to which women's occupational intentions are affected by the gender stereotypes associated with entrepreneurship. This may explain the positive interaction effect for women and the lack of a significant association between BAS-D and EI for males.

The proposed classification methods successfully predicted EI using BIS/BAS and FFM traits. Previous literature emphasized the importance of machine learning and classification algorithms in predicting FFM traits using Twitter texts (Pratama & Sarno, 2015) and identifying entrepreneurs using demographic variables (Montebruno et al., 2020). However, the prediction of EI using personality traits was thus far unexplored. Our results indicate that an RF algorithm using attribute selection can predict high EI using Extraversion, Openness, and BIS/BAS scores with an accuracy of 87.61%. Thus, when predicting whether an individual has high EI, the classifier is correct 87.61% of the time. As any accuracy above 50.00% is superior to random classification, we can thus confirm that personality is important for the prediction of EI. Furthermore, an LR using attribute selection can predict the rounded score (1 to 7) of an individual's EI with an accuracy of 32.00%. Overall, RF appeared to be the best classifier both in terms of accuracy and F1 score. This likely stems from its robustness against non-linear variables, noise, and overfitting.

Overall, BIS/BAS traits appear to have greater explanatory power than FFM traits in both OLS analysis and classification. This is implied by the variance explained by BIS/BAS in the regression analysis (0.242) and the independent predictive accuracy of BIS/BAS (83.33%) exceeding those of the FFM (0.178 and 80.34%, respectively). This comparison highlights the opportunity for further research on the BIS/BAS model in the field of personality and entrepreneurship, which is currently limited. Furthermore, the predictive accuracies and F1 scores of classifiers that used personality traits to predict binary EI were strictly higher than those of classifiers using control (demographic) variables for prediction (56.41%). This further supports the claim that personality is associated with EI.

6.2. Implications

Our study carries various implications for the ongoing debate of whether personality affects EI and for the fields of entrepreneurship, psychology, and machine learning. The OLS regression analysis presented in this paper identifies four personality traits associated with EI: BAS-D, BAS-FS, Conscientiousness, and Openness. This implies the possibility of trait paradigm as a determinant of EI. Furthermore, we identified multiple classification models that predict EI using personality traits at high levels of accuracy, suggesting a new direction for future research that combines machine learning, psychology, and economics. Therefore, we dispute Aldrich's (1999) claim that entrepreneurial personality research is at an empirical dead-end. Furthermore, contrary to Gartner's (1988) argument that an excessive number of personality traits are necessary to identify entrepreneurs, our classification analysis implies that features can be selected using wrappers, limiting the number of determining characteristics. This is reinforced by the 87.61% classification accuracy achieved using only six traits. Furthermore, the low accuracy of classifiers using control variables compared to a similar number of personality traits further emphasize the importance of personality traits as inputs for entrepreneurial decision-making.

In combination with further research, the traits and models explored in this paper could be used to identify specific target groups with high levels of EI, which could have important policy implications. For example, students who are likely to have greater entrepreneurial intentions could be identified and offered additional entrepreneurial education. Previous literature has identified a positive impact of entrepreneurial education on entrepreneurial outcomes (Matlay, 2008). Thus, providing support and education for individuals with high EI could increase the number of successful entrepreneurs, positively impacting the economy.

Lastly, while we could not statistically support our hypotheses concerning the moderating effect of gender, two significant interaction effects were identified. This encourages future research to investigate the role of gender in the relationship between personality traits and EI and could consequently add to the discussion of entrepreneurship as a gendered profession (Gupta et al., 2009).

6.3. Limitations & future research

The first limitation of our research is that the relationship between personality and EI remains unclear. While some evidence for an association between both personality measures and EI was found, we recommend additional research on both BIS/BAS and the FFM. An interesting research direction could include exploring the seemingly inconstant relationship between BAS-RR and EI (Geenen et al., 2016; Lerner et al., 2018a) or further exploring whether BIS/BAS's greater predictive and explanatory power remains when other datasets and methodologies are used. Furthermore, while

EI is a good predictor of entrepreneurship (Ajzen, 1991), it does not indicate whether the entrepreneur will be successful or not. Thus, the extent to which personality impacts an entrepreneur's success could be further investigated. This would allow the research results to carry greater implications for both economic and policy decision-making.

To continue, the method of data collection induces multiple limitations. Firstly, both the FFM and EI measures used only a selection of the questions proposed by the EIQ and the BFI. While EI's internal consistency (0.94) was excellent, Agreeableness's (0.49) was below the acceptable threshold. This means that the variable may not represent an individual's actual level of Agreeableness. Hence, we recommend that future surveys include a greater portion of the BFI to increase the FFM's internal consistency. Another limitation is that the sample's cross-sectional data does not allow us to establish causal relations. To improve the methodology, one could collect panel data that can be analysed and used to control for more factors. Lastly, as our sample consists only of university students, the external validity of our results is limited. This is because the sample is unlikely to represent the population, limiting the generalizability of its results. Future research could thus use a more diverse set of individuals in its data, creating a sample that better represents the general population.

A further limitation is the dataset's small sample. The dataset contains 150 students, which is relatively small for both OLS analysis and classification. Therefore, we suggest performing meta-data analysis on OLS regression results, as Zhao and Seibert (2006) did, allowing one to establish more robust relationships between personality traits and EI. Furthermore, we could use data merging to combine the data of studies with similar datasets to create a balanced dataset without using MWMOTE. A larger dataset size also increases the number of instances per category for categorical classification, which is currently only 21 ($150/7 \approx 21.4$). As a classifier's predictive power increases with the size and quality of the training set, its accuracy is also likely to increase (Figueroa et al., 2012). This presents an interesting direction for future research: researchers may predict EI using personality traits with great predictive accuracy using data merging.

Furthermore, our research limited the number of variables analysed to limit the length and scope of the study. However, multiple other personality measures, including self-confidence, personal attitude, and a proactive personality, have been associated with EI (Crant, 1996; Ferreira et al., 2012). Similarly, literature has associated EI with multiple demographic variables such as the level of education and the external environment (Crant, 1996; Khuong & An, 2016). Further research on these variables allows us to compare the predictive and explanatory abilities of demographic variables and personality traits more extensively, further contributing to the aforementioned debate on personality and EI. Lastly, the effect of personality on different types of entrepreneurship (i.e., necessity and opportunity entrepreneurship) could be further explored. This allows policymakers to target individuals that are likely to become productive opportunity entrepreneurs, increasing economic growth and innovation.

The last limitation is the classification algorithms used. Future research could include more complex classifiers such as Artificial Neural Networks (ANN), which do not restrict input variables and can learn more complex relations. AI is currently used in image recognition systems such as detecting bank fraud and in forecasting. However, no research on the forecasting of EI has been done thus far. Furthermore, the accuracy of the classifiers could be improved using feature engineering, algorithm tuning, and ensemble methods. Lastly, the number of cross-validation folds could be increased (currently, the standard ten folds are used). However, too many folds will decrease the accuracy by limiting the number of sample combinations.

6.4. Conclusion

The research question identified at the beginning of the paper was as follows: *"Are personality traits related to entrepreneurial intentions, and can these intentions be predicted?".* This question was split into two parts. First, an OLS regression analysis was conducted to identify whether personality traits are related to EI. Statistical support for hypotheses H1b, H1c, H2c, and H2e, was identified, implying that BAS-D, BAS-FS, Conscientiousness, and Openness are positively associated with EI. However, we conclude that additional research (i.e., meta data-analysis) on the association between personality and EI should be conducted to identify more consistent and significant regression results. While hypotheses 3a and 3b on the moderating effect of gender on EI and personality were rejected, there appeared to be some significant interaction effects between gender and personality. These effects should be further explored to add to the debate on entrepreneurship as a gendered profession.

Next, multiple classification algorithms explored whether personality traits can predict EI. The classifiers demonstrated that personality traits could successfully predict whether students have high versus low EI with an accuracy of 87.61%. Furthermore, a rounded EI score (1-7) can be predicted with an accuracy of 32.00%. Therefore, we conclude that EI can be predicted using personality traits but that further research could significantly improve this accuracy. Lastly, we suggest several directions for future research using meta-data analysis or data merging on datasets of personality traits, demographic variables, gender, EI, and different types of entrepreneurship.

7. References

1. Acs, Z. J., Arenius, P., Hay, M., & Minniti, M. (2005). Global entrepreneurship monitor: 2004 executive report. *Babson College and London Business School*.
2. Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179-211.
3. Akee, R., Copeland, W., Costello, E. J., & Simeonova, E. (2018). How does household income affect child personality traits and behaviors?. *American Economic Review*, *108*(3), 775-827.
4. Aldrich, H. (1999). *Organizations evolving*. Sage.
5. Audretsch, D. B., & Thurik, A. R. (1998). The knowledge society, entrepreneurship, and unemployment.
6. Azoulay, P., Jones, B. F., Kim, J. D., & Miranda, J. (2020). Age and high-growth entrepreneurship. *American Economic Review: Insights*, *2*(1), 65-82.
7. Bagheri, A., & Pihie, Z. A. L. (2014). The moderating role of gender in shaping entrepreneurial intentions: Implications for vocational guidance. *International Journal for Educational and Vocational Guidance*, *14*(3), 255-273.
8. Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, *44*(1), 1-26.
9. Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next?. *International Journal of Selection and assessment*, *9*(1-2), 9-30.
10. Barua, S., Islam, M. M., Yao, X., & Murase, K. (2012). MWMOTE--majority weighted minority oversampling technique for imbalanced data set learning. *IEEE Transactions on knowledge and data engineering*, *26*(2), 405-425.
11. Bernoster, I., Rietveld, C. A., Thurik, A. R., & Torrès, O. (2018). Overconfidence, optimism and entrepreneurship. *Sustainability*, *10*(7), 2233.
12. Brandstätter, H. (2011). Personality aspects of entrepreneurship: A look at five meta-analyses. *Personality and individual differences*, *51*(3), 222-230.
13. Brenner, S. L., Beauchaine, T. P., & Sylvers, P. D. (2005). A comparison of psychophysiological and self-report measures of BAS and BIS activation. *Psychophysiology*, *42*(1), 108-115.
14. Brockhaus, R. H. (1982). The psychology of the entrepreneur. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*.

15. Campbell-Sills, L., Liverant, G. I., & Brown, T. A. (2004). Psychometric evaluation of the behavioral inhibition/behavioral activation scales in a large sample of outpatients with anxiety and mood disorders. *Psychological assessment*, *16*(3), 244.
16. 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.
17. Caseras, X., Avila, C., & Torrubia, R. (2003). The measurement of individual differences in behavioural inhibition and behavioural activation systems: a comparison of personality scales. *Personality and individual differences*, *34*(6), 999-1013.
18. Cattell, R. B. (1946). Description and measurement of personality.
19. Cattell, R. B. (1947). Confirmation and clarification of primary personality factors. *Psychometrika*, *12*(3), 197-220.
20. Chartrand, C. (2016, April 5). *5 Reasons Why Personality at Work is Important!* AtmanCo. https://atmanco.com/blog/psychometrics/5-reasons-why-personality-at-work-is-more-important-than-you-think/
21. Ciavarella, M. A., Buchholtz, A. K., Riordan, C. M., Gatewood, R. D., & Stokes, G. S. (2004). The Big Five and venture survival: Is there a linkage?. *Journal of Business Venturing*, *19*(4), 465-483.
22. Cramer, J. S., Hartog, J., Jonker, N., & Van Praag, C. M. (2002). Low risk aversion encourages the choice for entrepreneurship: an empirical test of a truism. *Journal of Economic Behavior & Organization*, *48*(1), 29-36.
23. Cobb-Clark, D. A., & Schurer, S. (2012). The stability of big-five personality traits. *Economics Letters*, *115*(1), 11-15.
24. Costa, P. T., McCrae, R. R., & Holland, J. L. (1984). Personality and vocational interests in an adult sample. *Journal of Applied psychology*, *69*(3), 390.
25. Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory*. Odessa, FL: Psychological assessment resources.
26. Costa Jr, P. T., & McCrae, R. R. (1992). Revised NEO personality inventory (NEO-PI-R) and NEO five-factor (NEO-FFI) inventory professional manual. *Odessa, Fl: PAR*.
27. Crant, J. M. (1996). The proactive personality scale as a predictor of entrepreneurial intentions. *Journal of Small Business Management*, *34*, 42-49.
28. Cunningham, J. B., & Lischeron, J. (1991). Defining entrepreneurship. *Journal of Small Business Management*, *29*(1), 45-61.
29. Davidsson, P. (1995). Determinants of entrepreneurial intentions. In *RENT XI Workshop*.

30. Dietterich, T. (1995). Overfitting and undercomputing in machine learning. *ACM computing surveys (CSUR)*, *27*(3), 326-327.
31. Farrukh, M., Khan, A. A., Khan, M. S., Ramzani, S. R., & Soladoye, B. S. A. (2017). Entrepreneurial intentions: the role of family factors, personality traits and self-efficacy. *World Journal of Entrepreneurship, Management and Sustainable Development*.
32. Ferreira, J. J., Raposo, M. L., Rodrigues, R. G., Dinis, A., & do Paço, A. (2012). A model of entrepreneurial intention: An application of the psychological and behavioral approaches. *Journal of Small Business and Enterprise Development*.
33. Figueroa, R. L., Zeng-Treitler, Q., Kandula, S., & Ngo, L. H. (2012). Predicting sample size required for classification performance. *BMC Medical Informatics and Decision Making*, *12*(1), 1-10.
34. Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology*, *44*(3), 329.
35. Frank, E., & Hall, M. (2001, September). A simple approach to ordinal classification. In *European Conference on Machine Learning* (pp. 145-156). Springer, Berlin, Heidelberg.
36. Franken, I. H., & Muris, P. (2005). Individual differences in decision-making. *Personality and Individual Differences*, *39*(5), 991-998.
37. Gartner, W. B. (1988). “Who is an entrepreneur?” is the wrong question. *American Journal of Small Business*, *12*(4), 11-32.
38. Gartner, W. B. (1990). What are we talking about when we talk about entrepreneurship?. *Journal of Business venturing*, *5*(1), 15-28.
39. Geenen, N. Y., Urbig, D., Muehlfeld, K., van Witteloostuijn, A., & Gargalianou, V. (2016). BIS and BAS: Biobehaviorally rooted drivers of entrepreneurial intent. *Personality and Individual Differences*, *95*, 204-213.
40. Gray, J. A. (1972). The psychophysiological nature of introversion-extraversion: A modification of Eysenck's theory. *Biological Bases of Individual Behavior*, 182-205.
41. Gray, J. A. (1981). A critique of Eysenck’s theory of personality. In *A model for personality* (pp. 246-276). Springer, Berlin, Heidelberg.
42. Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition & emotion*, *4*(3), 269-288.
43. Gupta, V. K., Turban, D. B., Wasti, S. A., & Sikdar, A. (2009). The role of gender stereotypes in perceptions of entrepreneurs and intentions to become an entrepreneur. *Entrepreneurship Theory and Practice*, *33*(2), 397-417.
44. Hessels, J., Rietveld, C. A., Thurik, A. R., & Van der Zwan, P. (2018). Depression and entrepreneurial exit. *Academy of Management Perspectives*, *32*(3), 323-339.

45. Ismail, M., Khalid, S. A., Othman, M., Jusoff, H. K., Rahman, N. A., Kassim, K. M., & Zain, R. S. (2009). Entrepreneurial intention among Malaysian undergraduates. *International Journal of Business and Management*, *4*(10), 54-60.
46. John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues.
47. John, O. P., & Srivastava, S. (1999). *The Big-Five trait taxonomy: History, measurement, and theoretical perspectives*(Vol. 2, pp. 102-138). Berkeley: University of California.
48. Kautonen, T., Van Gelderen, M., & Fink, M. (2015). Robustness of the theory of planned behavior in predicting entrepreneurial intentions and actions. *Entrepreneurship Theory and Practice*, *39*(3), 655-674.
49. Khuong, M. N., & An, N. H. (2016). The factors affecting entrepreneurial intention of the students of Vietnam national university—a mediation analysis of perception toward entrepreneurship. *Journal of Economics, Business and Management*, *4*(2), 104-111.
50. Koestner, R., & Losier, G. F. (1996). Distinguishing reactive versus reflective autonomy. *Journal of Personality*, *64*(2), 465-494.
51. Kohavi, R. (1994, November). Feature subset selection as search with probabilistic estimates. In *AAAI fall symposium on relevance* (Vol. 224).
52. Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, *97*(1-2), 273-324.
53. Kohavi, R., & Sommerfield, D. (1995, August). Feature Subset Selection Using the Wrapper Method: Overfitting and Dynamic Search Space Topology. In *KDD* (pp. 192-197).
54. Komarraju, M., Karau, S. J., Schmeck, R. R., & Avdic, A. (2011). The Big Five personality traits, learning styles, and academic achievement. *Personality and Individual Differences*, *51*(4), 472-477.
55. Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology*, *49*(1), 1-49.
56. Krueger Jr, N. F., Reilly, M. D., & Carsrud, A. L. (2000). Competing models of entrepreneurial intentions. *Journal of business venturing*, *15*(5-6), 411-432.
57. Kuratko, D. F., Hornsby, J. S., & Naffziger, D. W. (1997). An examination of owner's goals in sustaining entrepreneurship. *Journal of Small Business Management*, *35*(1), 24.
58. Leone, L., & Russo, P. M. (2009). Components of the behavioral activation system and functional impulsivity: A test of discriminant hypotheses. *Journal of Research in Personality*, *43*(6), 1101-1104.

59. Lerner, D. A., Hatak, I., & Rauch, A. (2018a). Deep roots? Behavioral inhibition and behavioral activation system (BIS/BAS) sensitivity and entrepreneurship. *Journal of Business Venturing Insights*, *9*, 107-115.

60. Lerner, D. A., Hunt, R. A., & Dimov, D. (2018b). Action! Moving beyond the intendedly-rational logics of entrepreneurship. *Journal of Business Venturing*, *33*(1), 52-69.

61. Liñán, F., & Chen, Y. W. (2009). Development and cross–cultural application of a specific instrument to measure entrepreneurial intentions. *Entrepreneurship theory and practice*, *33*(3), 593-617.

62. Liñán, F., & Chen, Y. W. (2006). Testing the entrepreneurial intention model on a two-country sample.

63. Liñán, F., Rodríguez-Cohard, J. C., & Rueda-Cantuche, J. M. (2011). Factors affecting entrepreneurial intention levels: a role for education. *International Entrepreneurship and Management Journal*, *7*(2), 195-218.

64. Liu, B., Liu, J., & Hu, J. (2010). Person-organization fit, job satisfaction, and turnover intention: An empirical study in the Chinese public sector. *Social Behavior and Personality: an International Journal*, *38*(5), 615-625.

65. Mathews, C. H., & Moser, S. B. (1995). Family background and gender: Implications for interest in small firm ownership. *Entrepreneurship and Regional Development*, *7*(4), 365-378.

66. Matlay, H. (2008). The impact of entrepreneurship education on entrepreneurial outcomes. *Journal of Small Business and Enterprise Development*.

67. McDougall, W. (1929). The chemical theory of temperament applied to introversion and extroversion. *The Journal of Abnormal and Social Psychology*, *24*(3), 293.

68. Miller, B. K., Bell, J. D., Palmer, M., & Gonzalez, A. (2009). Predictors of entrepreneurial intentions: a quasi-experiment comparing students enrolled in introductory management and entrepreneurship classes. *Journal of Business & Entrepreneurship*, *21*(2).

69. Minniti, M., & Nardone, C. (2007). Being in someone else's shoes: the role of gender in nascent entrepreneurship. *Small business economics*, *28*(2), 223-238.

70. Montebruno, P., Bennett, R. J., Smith, H., & Van Lieshout, C. (2020). Machine learning classification of entrepreneurs in British historical census data. *Information Processing & Management*, *57*(3), 102210.

71. Moriano, J. A., Gorgievski, M., Laguna, M., Stephan, U., & Zarafshani, K. (2012). A cross-cultural approach to understanding entrepreneurial intention. *Journal of career development*, *39*(2), 162-185.

72. Mueller, P. (2006). Entrepreneurship in the region: Breeding ground for nascent entrepreneurs?. *Small Business Economics*, *27*(1), 41-58.
73. Papulová, Z., & Mokroš, M. (2007). Importance of managerial skills and knowledge in management for small entrepreneurs. *E-leader, Prague*, 1-8.
74. Pfeifer, S., Šarlija, N., & Zekić Sušac, M. (2016). Shaping the entrepreneurial mindset: Entrepreneurial intentions of business students in Croatia. *Journal of Small Business Management*, *54*(1), 102-117.
75. Pihie, Z. A. L., & Bagheri, A. (2013). Self-efficacy and entrepreneurial intention: The mediation effect of self-regulation. *Vocations and Learning*, *6*(3), 385-401.
76. Pratama, B. Y., & Sarno, R. (2015, November). Personality classification based on Twitter text using Naive Bayes, KNN and SVM. In *2015 International Conference on Data and Software Engineering (ICoDSE)* (pp. 170-174). IEEE.
77. Rauch, A., & Frese, M. (2007). Let's put the person back into entrepreneurship research: A meta-analysis on the relationship between business owners' personality traits, business creation, and success. *European Journal of work and organizational psychology*, *16*(4), 353-385.
78. Rey-Martí, A., Porcar, A. T., & Mas-Tur, A. (2015). Linking female entrepreneurs' motivation to business survival. *Journal of business research*, *68*(4), 810-814.
79. Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology*, *58*(1), 103-139.
80. Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological bulletin*, *126*(1), 3.
81. Robledo, J. L. R., Arán, M. V., Sanchez, V. M., & Molina, M. Á. R. (2015). The moderating role of gender on entrepreneurial intentions: A TPB perspective. *Intangible capital*, *11*(1), 92-117.
82. Romer, D., Betancourt, L., Giannetta, J. M., Brodsky, N. L., Farah, M., & Hurt, H. (2009). Executive cognitive functions and impulsivity as correlates of risk taking and problem behavior in preadolescents. *Neuropsychologia*, *47*(13), 2916-2926.
83. Sandhu, M. S., Sidique, S. F., & Riaz, S. (2011). Entrepreneurship barriers and entrepreneurial inclination among Malaysian postgraduate students. *International Journal of Entrepreneurial Behavior and Research*.
84. Scheres, A., & Sanfey, A. G. (2006). Individual differences in decision making: Drive and reward responsiveness affect strategic bargaining in economic games. *Behavioral and brain functions*, *2*(1), 1-8.

85. Schneider, B. (1987). The people make the place. *Personnel psychology*, *40*(3), 437-453.
86. Schneider, T. R. (2004). The role of neuroticism on psychological and physiological stress responses. *Journal of Experimental Social Psychology*, *40*(6), 795-804.
87. Singh, G., & DeNoble, A. (2003). Views on self-employment and personality: An exploratory study. *Journal of Developmental Entrepreneurship*, *8*(3), 265.
88. Stewart, W. H., & Roth, P. L. (2007). A meta-analysis of achievement motivation differences between entrepreneurs and managers. *Journal of small business management*, *45*(4), 401-421.
89. Suls, J. (2001). Affect, stress, and personality. *Handbook of affect and social cognition*, 392-409.
90. Tupes, E.C., & Christal, R.E. (1961). *Recurrent personality factors based on trait ratings* (ASD-TR-61-97). Lackland Air Force Base, TX: Aeronautical Systems Division, Personnel Laboratory.
91. Wennekers, S., & Thurik, R. (1999). Linking entrepreneurship and economic growth. *Small Business Economics*, *13*(1), 27-56.
92. Hatak, I., Harms, R., & Fink, M. (2015). Age, job identification, and entrepreneurial intention. *Journal of Managerial Psychology*.
93. Vamvaka, V., Stoforos, C., Palaskas, T., & Botsaris, C. (2020). Attitude toward entrepreneurship, perceived behavioral control, and entrepreneurial intention: dimensionality, structural relationships, and gender differences. *Journal of Innovation and Entrepreneurship*, *9*(1), 5.
94. Van der Zwan, P., Thurik, R., & Grilo, I. (2010). The entrepreneurial ladder and its determinants. *Applied Economics*, *42*(17), 2183-2191.
95. Verheul, I., Thurik, R., Grilo, I., & Van der Zwan, P. (2012). Explaining preferences and actual involvement in self-employment: Gender and the entrepreneurial personality. *Journal of economic psychology*, *33*(2), 325-341.
96. Vukasović, T., & Bratko, D. (2015). Heritability of personality: A meta-analysis of behavior genetic studies. *Psychological bulletin*, *141*(4), 769.
97. Wang, C. L. (2008). Entrepreneurial orientation, learning orientation, and firm performance. *Entrepreneurship theory and practice*, *32*(4), 635-657.
98. Warr, P., & Inceoglu, I. (2012). Job engagement, job satisfaction, and contrasting associations with person–job fit. *Journal of Occupational Health Psychology*, *17*(2), 129.
99. Willem, L., Bijttebier, P., & Claes, L. (2010). Reactive and self-regulatory temperament dimensions in relation to alcohol use in adolescence. *Addictive behaviors*, *35*(11), 1029-1035.
100. Yen, T. (2021, May 20). *How to Make Better Decisions About Your Career*. Harvard Business Review. https://hbr.org/2021/05/how-to-make-better-decisions-about-your-career
101. Zhao, F. (2005). Exploring the synergy between entrepreneurship and innovation. *International Journal of Entrepreneurial Behavior & Research*.

102. Zhao, H., & Seibert, S. E. (2006). The big five personality dimensions and entrepreneurial status: A meta-analytical review. *Journal of Applied Psychology*, *91*(2), 259.

103. Zhao, H., Seibert, S. E., & Lumpkin, G. T. (2010). The relationship of personality to entrepreneurial intentions and performance: A meta-analytic review. *Journal of Management*, *36*(2), 381-404.

8. Appendix

8.1. A1: Conceptual models of hypotheses

Diagram 1a: *Conceptual Model (1) of expected relationships (BIS/BAS)*

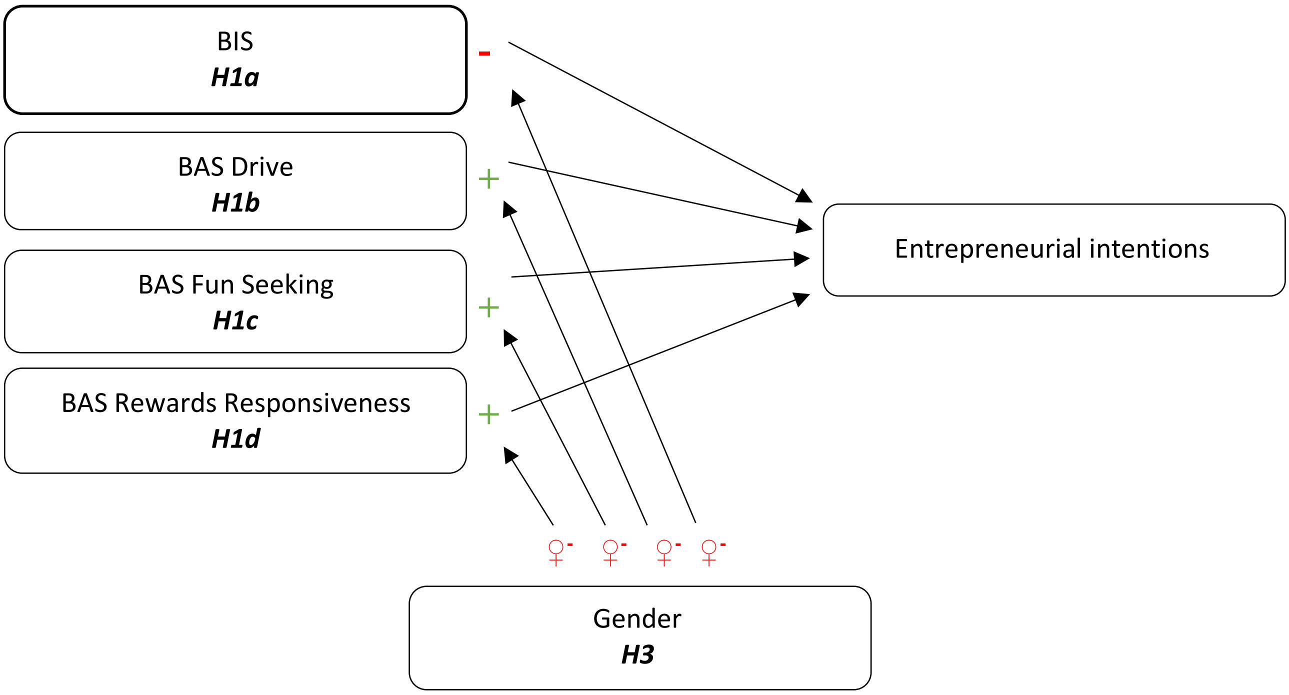
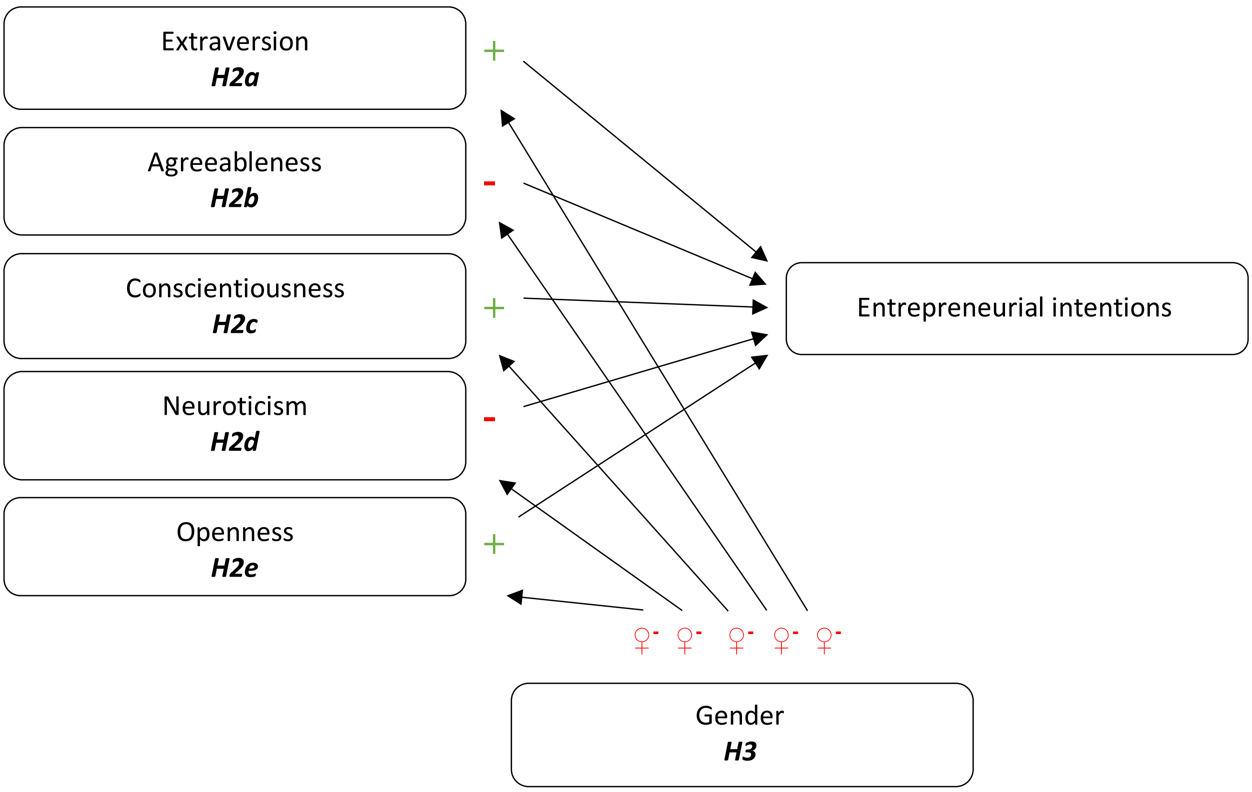
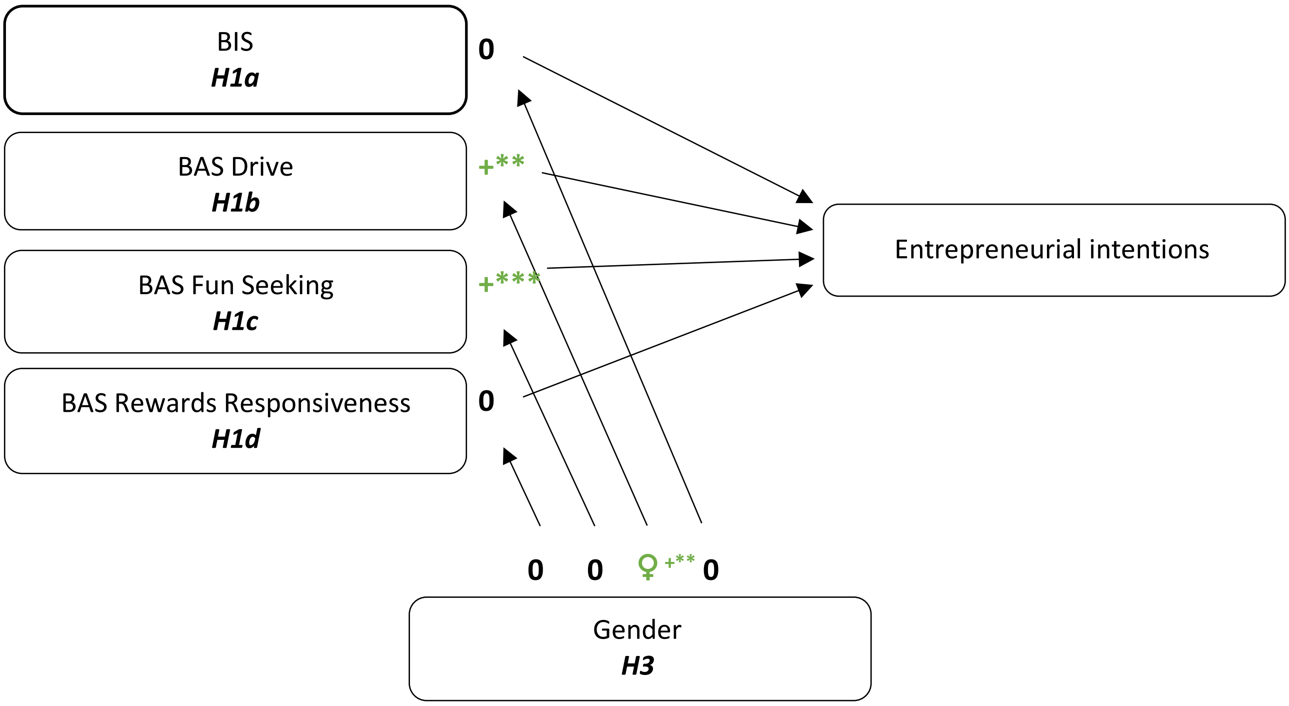


Diagram 1b: *Conceptual Model (2) of expected relationships (FFM)*



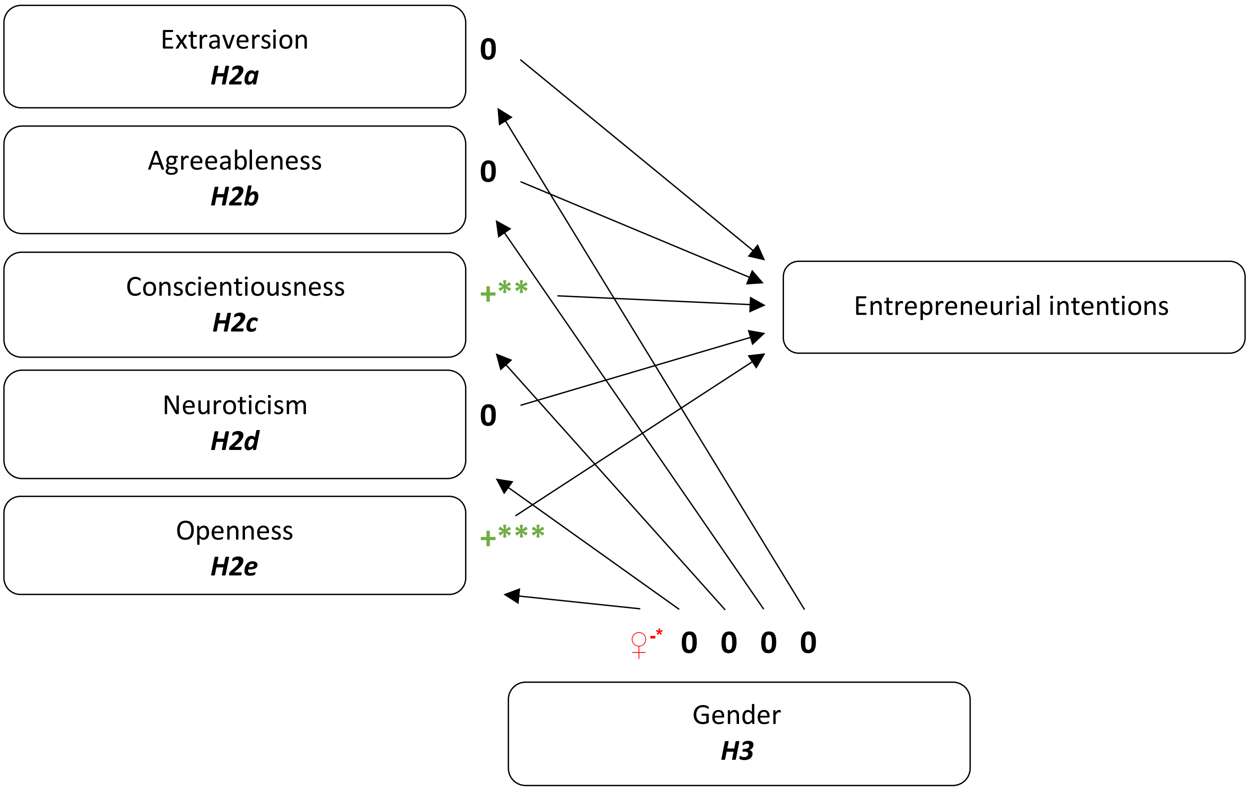
8.2. A2: Overview of regression analysis results

Diagram 2a: *Conceptual Model (3) of found associations (BIS/BAS)*



Note. * = p<0.1, ** = p<0.05, *** = p<0.01.

Diagram 2b: *Conceptual Model (4) of found associations (FFM)*



Note. * = p<0.1, ** = p<0.05, *** = p<0.01.

8.3. A3: Control variable classification

Table 6a: *Attribute selection for each classifier and dataset for control variables*

Classifier	Binary	Balanced Binary	Ordinal
Logistic Regression	Age	Ent. Parent	Ent. Parent
	Age2		
	Ent. Parent		
Naïve Bayes	Age	Ent. Parent	Age
			Ent. Parent
Random forest	NA*	Ent. Parent	Ent. Parent

Note. For the ordinal dataset, the attribute selection took place with the ordinal class classifier version of each of the classification techniques; NA* indicates that the wrapper did not select any variables during the wrapping process.

Table 6b: *Classification results for control variables*

Classifier		Logistic Regression	Naïve Bayes	Random forest
Binary	Accuracy	79.33%	76.67%	70.67%
	F1	0.714	0.715	0.662
Binary*	Accuracy	79.33%	78.67%	NA***
	F1	0.714	0.710	NA***
Balanced Binary	Accuracy	52.14%	50.00%	56.41%
	F1	0.521	0.500	0.563
Balanced Binary *	Accuracy	55.13%	55.13%	55.13%
	F1	0.514	0.514	0.514
Ordinal	Accuracy	30.00%	30.00%	23.33%
	F1	NA**	NA**	0.226
Ordinal*	Accuracy	33.33%	36.00%	33.33%
	F1	NA**	NA**	NA**

Note. * indicates that only attributes selected in the WrapperSubsetEval method are included. For the ordinal dataset, the attribute selection took place with the ordinal class classifier version of each of the classification techniques; NA** indicates that one or more of the components of the F1 score cannot be calculated. Therefore, the F1 score cannot be calculated and is not displayed; NA*** indicates that no variables were selected during the wrapping process. Therefore, neither the accuracy nor the F1 score can be calculated for these models.

8.4. A4: BIS/BAS classification

Table 7a: *Attribute selection for each classifier and dataset for BIS/BAS*

Classifier	Binary	Balanced Binary	Ordinal
Logistic Regression	BAS-D	BAS-D	BAS-FS
		BAS-FS	
		BAS-RR	
Naïve Bayes	BAS-D	BAS-D	NA*
	BAS-FS	BAS-FS	
	BAS-RR	BAS-RR	
Random forest	BAS-RR	BAS-D	NA*
		BAS-FS	
		BAS-RR	
		BIS	

Note. For the ordinal dataset, the attribute selection took place with the ordinal class classifier version of each of the classification techniques; NA* indicates that the wrapper did not select any variables during the wrapping process.

Table 7b: *Classification results for BIS/BAS variables*

Classifier		Logistic Regression	Naïve Bayes	Random forest
Binary	Accuracy	78.67%	80.00%	78.67%
	F1	0.764	0.795	0.768
Binary*	Accuracy	80.67%	80.67%	80.67%
	F1	0.784	0.803	0.757
Balanced Binary	Accuracy	72.65%	75.21%	83.33%
	F1	0.726	0.752	0.833
Balanced Binary *	Accuracy	73.93%	74.79%	83.33%
	F1	0.739	0.748	0.833
Ordinal	Accuracy	20.00%	22.00%	21.33%
	F1	0.165	0.189	0.207
Ordinal	Accuracy	27.33%	NA***	NA***
	F1	NA**	NA***	NA***

Note. * indicates that only attributes selected by the WrapperSubsetEval method are included. For the ordinal dataset, the attribute selection took place with the ordinal class classifier version of each of the classification techniques; NA** indicates that one or more of the components of the F1 score cannot be calculated. Therefore, the F1 score cannot be calculated and is not displayed; NA*** indicates that no variables were selected during the wrapping process. Therefore, neither the accuracy nor the F1 score can be calculated for these models.

8.5. A5: FFM classification

Table 8a: *Attribute selection for each classifier and dataset for FFM.*

Classifier	Binary	Balanced Binary	Ordinal
Logistic Regression	NA*	Neuroticism	Extraversion
		Openness	Agreeableness
Naïve Bayes	NA*	Extraversion	Extraversion
		Conscientiousness	Neuroticism
		Neuroticism	
		Openness	
Random forest	NA*	Extraversion	Conscientiousness
		Agreeableness	
		Neuroticism	
		Openness	

Note. For the ordinal dataset, the attribute selection took place with the ordinal class classifier version of each of the classification techniques; NA* indicates that the wrapper did not select any variables during the wrapping process.

Table 8b: *Classification results for FFM variables*

Classifier		Logistic Regression	Naïve Bayes	Random forest
Binary	Accuracy	76.67%	74.67%	74.67%
	F1	0.698	0.702	0.686
Binary*	Accuracy	NA***	NA***	NA***
	F1	NA***	NA***	NA***
Balanced Binary	Accuracy	65.81%	71.79%	80.34%
	F1	0.658	0.716	0.803
Balanced Binary *	Accuracy	67.09%	70.94%	79.06%
	F1	0.669	0.708	0.790
Ordinal	Accuracy	27.33%	22.67%	25.33%
	F1	NA**	0.189	0.241
Ordinal	Accuracy	30.67%	31.33%	26.67%
	F1	NA**	NA**	0.218

Note. * indicates that only attributes selected by the WrapperSubsetEval method are included. For the ordinal dataset, the attribute selection took place with the ordinal class classifier version of each of the classification techniques; NA** indicates that one or more of the components of the F1 score cannot be calculated. Therefore, the F1 score cannot be calculated and is not displayed; NA*** indicates that no variables were selected during the wrapping process. Therefore, neither the accuracy nor the F1 score can be calculated for these models.

8.6. A6: R-Code for MWMOTE

```
# Open imbalance package
library(imbalance)

# Open the unbalanced data set
UnbalancedData<- read.csv(file.choose(), header =T)

# Create new instances where students have high EI using MWMOTE
newCases<- mwmote(UnbalancedData, numInstances = 84, classAttr = "EntInt")

# Combine the new and old data into one data set
newData = rbind(UnbalancedData, newCases)

# Convert the new data set into a csv file
write.csv(newData, "WEKAset.csv", row.names = FALSE)
```

Note that "EntInt" in the third line of code is short for entrepreneurial intentions, the EI variable.

8.7. A7: Breusch-Pagan test for heteroskedasticity

The Breusch-Pagan test indicates whether linear regression models suffer from heteroscedasticity. The null hypothesis states that there is a constant variance of errors (homoscedasticity). Thus, when we rejected the null hypothesis at the 5% level ($p < 0.05$), the model likely suffers from heteroscedasticity. The results of the Breusch-Pagan test (Table 9) indicate that most regression Models (2, 4, 5, and 7) suffer from severe heteroskedasticity, while Models 3 and 4 suffer from slight heteroscedasticity ($p < 0.1$). Therefore, robust SEs are used in regressions.

Table 9: *Breusch-Pagan*

	(Model 1)	(Model 2)	(Model 3)	(Model 4)	(Model 5)	(Model 6)	(Model 7)
Chi	0.15	9.89	2.99	7.67	11.15	3.29	9.73
Prob > chi2	0.701	0.002	0.084	0.006	0.001	0.070	0.002
H_o rejected	No	Yes	No	Yes	Yes	No	Yes

Note. The column number indicates the model of the OLS regression that the Breusch-Pagan test has been applied to. This corresponds to the model number in tables 3a and 3b in section 5.1.

8.8. A8: VIF analysis

VIF analysis indicates whether the independent variables in linear regression models suffer from multicollinearity. When the VIF value exceeds 5, multicollinearity is likely to create poorly predicted coefficients and inaccurate p-values that cannot be interpreted. Table 10 indicates that neither the BIS/BAS nor the FFM regression suffers from severe multicollinearity. Therefore, the models of personality can be used in regression analysis, and we can assume that the multicollinearity assumption is not violated.

Table 10: *VIF calculations*

Variable	**VIF**
*Controls**	
Gender	1.05
Ent. Parent	1.04
Job	1.04
Age	1.01
Mean	1.04
BIS/BAS	
BAS-D	1.74
BAS-FS	1.53
BAS-RR	1.21
BIS	1.21
Mean	1.42
FFM	
Extraversion	1.07
Agreeableness	1.02
Conscientiousness	1.03
Neuroticism	1.02
Openness	1.06
Mean	1.04

Note. *For the Controls, Age^2 was not included in VIF calculations because it adds structural multicollinearity to the VIF, while we are only interested in data multicollinearity.