Freie Universität Berlin
Fachbereich Erziehungswissenschaft und Psychologie

Looking for c(l)ues.

How visual cues can help predict personality traits in video interviews

Dissertation
zur Erlangung des akademischen Grades
Doktor der Philosophie (Dr. phil.) in Psychologie

vorgelegt von
Richard T. Justenhoven
Berlin, 2022

First Advisor (Erstgutachter):

Prof. Dr. Stefan Krumm

Secondary Advisor (Zweitgutachter):

Jun.-Prof. Dr. Philipp Schäpers

Date of Disputation: 6th February 2023

I dedicate this thesis to my wonderful wife Lena,
who taught me to believe in what I am capable of, and
without whom this thesis would not have been possible.

Acknowledgements

A very special thanks goes to **Prof. Dr. Stefan Krumm** and **Dr. Achim Preuß** who have been the two people that put their trust and support into me to enable me in starting this journey. Both have been amazing providers of ideas, unconventional thoughts and approaches – but even more helpful a guide on what to focus on, and what not. Thank you to **Dr. Katharina Lochner**, the person who inspired me to follow my passion for research and consider this opportunity as well as to **Prof. Dr. Philipp Schäpers** to review this body of work and take up the responsibility as secondary adviser.

I had the chance to complete this work next to continuing to pursue my career within AON, for which I am grateful to **Dr. Nicholas Martin**, **Jouko van Aggelen** and **Tarandeep Singh** for their understanding and patience.

A very huge thanks to my esteemed colleagues **Maximilian Jansen,** and **Eoin O'Callaghan,** who went the extra mile in ensuring I had the time to focus on (and finish) this research work and acted as a sound board to initial ideas and thoughts around this endeavour.

A big thank goes to my entire family who ensured I had an environment to concentrate, thrive and work on my thesis, who challenged me when needed and motivated me with words, coffee, dinner, and other bribes to stay on track.

Finally, I am grateful to have friends that ensured I don't lose touch with the real world; and offered me balance through rowing and work in the firefighting department. Special thanks to **Hanns-Christian Neumann**, and **Andreas Siebert** to lead such excellent teams that I am proud to be a part of.

Abstract

There is an ongoing trend that asynchronous video interviews are used more and more frequently for their efficiency gain (Brenner, 2019), especially in large scale selection processes (Brandt, Justenhoven & Schöffel, 2020). Visual cues that are present during those video recordings are not yet systematically processed and used, which is a miss under the argument of further efficiency gain (e.g. to measure personality traits). However, in a first step, a framework as well as a systematic visual cue analyses needs to be completed to establish the available data source that can – in a second step – be further used in an automatic scoring process.

The purpose of the present study exactly that first step: to outline an approach to capture, categorize and systematically process visual cues to then link them to personality traits which are captured in various other forms as well. As an approach to this topic, the work from Gosling and colleagues (2005) is leveraged and with it Brunswik's lens model (1956). Ultimately, and postulated as a research question, the aim of this body of research is to find functional achievement between self-rated and observer rated personality traits using the visual cues as elements of the lens.

The body of research is structured in three steps. In step 1 visual cues present in research are catalogued, enriched with visual cues captured through various studies with asynchronous video interview data and categorized in five categories: Face, Body, Appearance, Media Properties, and Environment. In step 2 a visual cue inventory is developed that allows a manual systematic cue coding process of the 236 visual cues that are used in this work. In step 3, a dataset with n = 99 participants is generated that includes coding for all of the visual cues, as well as self and observer ratings on the video respondee' s personality traits.

Contrary to the hypotheses, however, little evidence is found that suggest visual cues can be linked both to self-ratings and observer ratings of personality traits. The cues seem to be either valid (i.e. linked to self-ratings) or used (i.e. linked to observer ratings) but generally the results show a very confound picture.

Given the present results, it is not recommended to proceed further with the approach to leverage visual cues as a predictor for personality traits in asynchronous video interviews.

Zusammenfassung

Es gibt einen anhaltenden Trend, dass asynchrone Videointerviews wegen ihres Effizienzgewinns immer häufiger eingesetzt werden (Brenner, 2019), insbesondere in groß angelegten Auswahlverfahren (Brandt, Justenhoven & Schöffel, 2020). Visuelle Hinweisreize, die während dieser Videoaufnahmen vorhanden sind, werden noch nicht systematisch verarbeitet und verwendet, was unter dem Argument der weiteren Effizienzsteigerung (z. B. zur Messung von Persönlichkeitsmerkmalen) ein Versäumnis ist. In einem ersten Schritt muss jedoch ein Rahmenwerk sowie eine systematische Analyse der visuellen Hinweise geschaffen werden, um die verfügbare Datenquelle zu ermitteln, die in einem zweiten Schritt in einem automatischen Scoring-Prozess weiterverwendet werden kann.

Das Ziel der vorliegenden Studie ist genau dieser erste Schritt: einen Ansatz zur Erfassung, Kategorisierung und systematischen Verarbeitung visueller Hinweise zu skizzieren, um diese dann mit Persönlichkeitsmerkmalen zu verknüpfen, die auch in verschiedenen anderen Formen erfasst werden. Als Grundlage zu diesem Thema wird die Arbeit von Gosling und Kollegen (2005) genutzt und damit das Linsenmodell von Brunswik (1956). Letztlich, und als Forschungsfrage postuliert, ist das Ziel der vorliegenden Forschung, die funktionale Leistung zwischen selbst- und beobachterbewerteten Persönlichkeitsmerkmalen – unter Verwendung der visuellen Hinweise als Elemente der Linse – zu finden.

Die Forschungsarbeit ist in drei Schritte gegliedert. In Schritt 1 werden die in der Forschung vorhandenen visuellen Anhaltspunkte katalogisiert, mit visuellen Anhaltspunkten angereichert, die in verschiedenen Studien mit asynchronen Videointerviewdaten erfasst wurden, und in fünf Kategorien kategorisiert: Gesicht, Körper, Erscheinungsbild, Medieneigenschaften und Umgebung. In Schritt 2 wird ein Inventar visueller Hinweisreize entwickelt, das einen manuellen systematischen Kodierungsprozess der 236 visuellen Hinweisreize ermöglicht, die in dieser Arbeit verwendet werden. In Schritt 3 wird ein Datensatz mit n = 99 Teilnehmern generiert, der die Kodierung aller visuellen Hinweisreize sowie Selbst- und Beobachtereinschätzungen zu den Persönlichkeitsmerkmalen des Video-Respondenten enthält.

Im Gegensatz zu den Hypothesen finden sich jedoch nur wenige Hinweise darauf, dass visuelle Hinweisreize sowohl mit Selbst- als auch mit Fremdeinschätzungen von Persönlichkeitsmerkmalen verknüpft werden können. Die Hinweisreize scheinen entweder gültig zu sein (d. h. mit Selbsteinschätzungen korrelierend) oder verwendet zu werden (d. h. mit Beobachtereinschätzungen korrelierend), aber im Allgemeinen zeigen die Ergebnisse ein sehr diffuses Bild auf.

In Anbetracht der vorliegenden Ergebnisse wird nicht empfohlen, den Ansatz weiter zu verfolgen, visuelle Hinweise als Prädiktor für Persönlichkeitsmerkmale in asynchronen Videointerviews zu nutzen.

Content

List of figures

List of tables

Chapter 1: Introduction

Problem Statement

The world of work is changing more and more toward being digital and virtual. Working arrangements change and, while causing challenges in selection, recruitment and development, it also allows talent to consider a broader range of employers, to follow their dreams outside location boundaries and for companies to recruit a more diverse workforce. The latest report of the World Economic Forum projects that, by 2025, "on average, 6% of workers are expected to be fully displaced" (p. 8, Schwab & Zahidi, 2020). Furthermore, it states that approximately 44% of the skills that current employees have in order to perform their daily tasks will be obsolete (and will be replaced by other skills or abilities). The report shows that over half (55%) of the companies surveyed expect to essentially change their value chain by 2025 due to new technologies, such as quantum computing, artificial intelligence, cloud computing and robotics (Schwab & Zahidi, 2020).

This does lead to the question: to what extent will the use of technology continue to change in recruitment and selection processes, allowing global and virtual communication (and more?) with potential candidates, new hires and employees? Already, video interview technology is used heavily in current large-scale selection, as well as development processes; it is expected that this trend will continue (Brandt, Justenhoven & Schöffel, 2020). Numbers vary for markets and regions, though most reports show an increase of usage of video interviews in selection processes of between 67% (Robert Walters Group, n.d.) and 87% (Golden, 2020) compared to the respective previous year for the last two years. The Covid crisis is named as a core driver for this direction as it drove companies to seek alternative routes to face-2-face meetings and interviews.

In the early stages of video interview tools, it took a while for the industry to gain momentum. HireVue, a video-interviewing company that was among the first to offer pre-recording videos as interview format shipped webcams to users; a hardware that is required but not used by most users at the turn of the century (VidCruiter, n.d.).

As with other digital media and devices, video interviews have seen an increasingly fast development of new capabilities which not only affect overall usage and adoption, but also affect how people interact with the technology and how they perceive its use (especially in a context as critical as personnel selection). An example of this is provided by Bartram (2000) predicting that "by the middle of this decade, we will see domestic digital TV with built-in cameras being used as video phones as part of their role as general purpose multi-media entertainment and information centres" (p. 267). It is interesting to note, that even though Bartram (2000) mentions the possibility of using a "(…) digital mobile phone or TV internet browser (…)" (p. 267) to fill out questionnaires and schedule an interview, his prediction was strongly influenced by a TV being the core piece of technology in a household. While the general notion of multi-purpose devices being the central point of interactions with various media have come true, most recent developments have favoured smaller and personal mobile devices, such as the Apple iPhone which was first released in 2007 (Apple Inc., 2007) and its tablet counterpart the iPad, first released in 2010 (Apple Inc., 2010). This is not only an interesting note on technological development, but also affects the way technology-mediated interviews are conducted, with mobile devices and personal computers offering much greater flexibility for applicants to choose the location of their interview. This greatly increases the possible variance in the kinds of backgrounds visible to interviewers. Additionally, video calls have become an almost everyday occurrence for many people, making video interviews a much less novel and out-of-the-ordinary situation. This could make video interviews less stressful for an applicant in 2022, compared to someone in the early 2000s, and allow them to behave more naturally in expressing their personality. It seems reasonable to assume that, in addition to mobility and familiarity, the interviewer representation could also affect applicant behaviour and contribute to very different situational characteristics when comparing TV-based video interview setups with mobile devices or personal computers. For example, Straus, Miles and Levesque (2001) mention using a 48-inch TV screen in a conference room to conduct videoconference interviews in their study on the effects of different interview types on applicant judgements. An interviewer presented in such a setup might appear considerably more imposing than someone viewed on a 5-to-7 inch smartphone or a 13-to-16 inch laptop screen in a home environment.

In practice, citing research that is almost 20 years old is not unthinkable if the source provides relevant data and insights or it is of a foundational nature and relevance in its field. However, as the above examples illustrate, direct applicability of methods, results and conclusions to current research cannot always be assumed. This further highlights the need for frameworks and theories on factors influencing interview outcomes that are flexible enough to be applicable to different situations, iterations of technological development and also designed to be open to adaptation and further development.

One, if not the major driver, for the use of video interviews is the efficiency gain of this method over costly and logistic-intense face-to-face interviews (Brenner, 2019).

In the same manner, additional technologies are developed and implemented that further increase the efficiency gain, especially the rating process. Examples for those are the implementation of automatic facial action unit programs, based on Ekman's study of emotions and his development of the six-emotion system. Or the implementation of spectral audio characteristics that extract features, such as pitch, volume and gaps in speech, to be used for feature-trained learning algorithms and modelled to predict personality traits. (Leutner, Akhtar & Chamorro-Premuzic, 2022)

Advanced data science models and machine-learning algorithms are central to these developments. They are used to incorporate millions of individual datapoints, even those that are generated with little or no additional effort, allowing impressive scalability opportunities while reducing the individual effort and thereby improving the efficiency of these processes to unprecedented highs (Javed & Brishti, 2020; Leutner, Akhtar & Chamorro-Premuzic, 2022).

A purely data-driven approach derives findings from the data itself without having a theoretical framework that backs the findings (Burkov, 2019). Such approaches tend to prioritise predictive validity in domain-specific use cases over consistency across different applications (Burkov, 2019). As such, a purely data-driven approach can bear considerable risks if done without a wider framework. Deriving insights based on available data can lead to overfitted models, lack of explainability or lack contextualisation. This, in turn, makes it more problematic to generalise findings over and above defined samples and even more so making it hard to ensure overall fairness towards users of these models (Brandt, Justenhoven & Schöffel, 2020; Burkov, 2019; Leutner, Akhtar & Chamorro-Premuzic, 2022).

In the interest of facilitating research on the psychometric properties of different models and approaches in this context, as well as being able to provide evidence-based recommendations for practitioners, it is important to ensure these newly-developed methods are aligned with the key principles of psychometric measurement. One of which is a theory-driven approach where hypotheses are defined a-priori and data is processed to test the theory.

At present, it is clear that there are benefits of automatic rating for video interviews in selection processes (Brandt, Justenhoven & Schöffel, 2020). However, there are also vast areas that miss foundational research to enable a more holistic and psychometrically-sound automatic rating of video interviews. This is where the present work aims to build bridges. Such research becomes even more relevant, as automated rating tools see increasing adoption while regulatory frameworks and guidance on the development and use of machine learning in personnel selection – such as a recent New York City law on automated employment decision tools (2021) and ethics guidelines for trustworthy AI by the European Commission (2019) are also on the rise.

Firstly, interview raters rely heavily on visual information, which can constitute both a valid source of information and a bias that adds noise and, with that, errors (Gorden, 1998). However, there is little research to date that provides a full overview of the specific visual information that is available during video interviews – specifically for those that are highly structured, such as asynchronous video interviews. Given their structured nature and the fact that they can be described by a preselected setting with planning regarding time and location, it would be especially interesting to investigate the value of understanding visual information. Something not possible though without understanding, categorising and validating potential available visual information first.

Secondly, there are methodical knowledge gaps on how available visual information in (asynchronous) video interviews can be captured and processed in a structural way that is in line with standard psychometric approaches and quality criteria.

Thirdly, and this is where psychometric principles are already being applied, the visual information is to be linked to established constructs that are to be measured. Clearly one of the most frequently-used constructs measured during video interviews in selection processes is personality.

The above chain of research and insights, that is needed to enable a more holistic automatic rating of personality traits in video interviews, is somewhat too ambitious to tackle in a single research initiative. However, this work is starting to shed light on how visual information from asynchronous video interviews in the setting of selection processes can be structured and processed and how it relates to the personality (factors) of the interviewees, both from their self-perception as well as how it is judged by observers.

Work Structure

The following three chapters aim to help answer the previously-posed questions. Chapter 2 will be laying the theoretical foundation for the research topic and upcoming studies. Among others, relevant theoretical models of trait judgements in interviews relevant to this work are presented as well as related information to employment interviews and, more specifically, to video-based interviews. Also, the chapter provides a short overview about different factors that can influence the decision-making process during interviews, both from the interviewer's as well as the interviewee's side.

Chapter 3 focuses on the different frameworks and methods that are used and leveraged in the studies of this work. Among those, the IPIP and ADEPT-15 models and respective questionnaires are introduced, as well as the lens model by Brunswik and the Personal Living Space Cue Inventory by Gosling.

Chapter 4 will dive into the different steps of the studies conducted over the course of this research initiative, highlighting the collection of visual information through literature review and data collection studies and how the gathered information is turned into a classification and inventory for future use. The chapter explains the data collection and the main study's results to answer the overarching research questions.

Chapter 5 closes this work and will put the findings and work into a broader context. A discussion will highlight to what extent the research questions have been answered and which areas will remain open. This, in turn, opens up additional avenues for future research, but also a critical review of how the research was carried out.

Chapter 2: Theoretical Background

The following sections are intended to provide an overview of some of the key research fields, methods, theories and models that the present work is based on. Following a brief introduction to employment interviews and their role in personnel selection processes, several theoretical models of trait judgements in interviews relevant to this work are presented. As with large parts of this work's research design, the focus is on trait judgements operationalised through self-other agreement, which is why the Realistic Accuracy Model by Funder (1995), the Self-Other Knowledge Asymmetry Model by Vazire (2010) and the Trait Reputation Identity Model by McAbee and Connelly (2016) were deemed to be especially relevant.

Following these models, a range of factors influencing interview processes and outcomes both on the side of the interviewer and the interviewee are presented. This includes effects of initial impressions, impression management by candidates and candidate characteristics. Thin slices are briefly outlined as relevant methodological concept related to interviewers forming early impressions of candidates that is also used in this work's research. The overall relevance of research in this area is highlighted by the possible consequences of interview outcomes, including the costs of false hiring decisions. Influences on interview outcomes are further explicated following the differentiation of four moderators of accurate interpersonal judgements (good judge, good target, good trait and good information) proposed by Funder (1995) as part of the Realistic Accuracy Model.

As this work focuses on video interviews, a general introduction to this format is provided. This is supported by theories on media richness and synchrony which can help explain differences between interview formats with varying degrees of technological mediation. A two-factor classification for technologically-mediated interviews is proposed. Asynchronous video interviews are the most relevant format of technologically-mediated interviews in the context of this work and are thus presented in more detail, including automatic evaluations and their perception by both interviewees and interviewers and implications for bias and fairness.

Finally, a brief introduction to the use of visual information in asynchronous video interviews and some general expectations for different traits based on previous research, as well as the research questions guiding this work are presented.

Introduction and Context

The employment interview is one of the most widely used, continually researched and often referred to be among the best perceived personnel selection tool (Brenner, 2019; Brenner, Ortner & Fay, 2016; Huffcutt, Van Iddekinge & Roth, 2011; Huffcutt & Youngcourt, 2007; Levashina et al., 2014; Schmidt, Oh & Shaffer, 2016, Torres & Gregory, 2018). While traditionally defined as face-to-face interaction, technological development and adoption have increased the number of possible technologically-mediated interview variations (Huffcutt & Youngcourt, 2007; Levashina et al., 2014; Torres & Gregory, 2018). Brenner (2019) provides a recent and thorough overview of different interview variations and components across various digital, non-digital and hybrid forms based on Campion, Palmer and Campion's (1997) and Levashina and colleagues' (2014) combined 18 components of interview structure, as well as Chapman and Zweig's (2005) four dimensions of interview structure. In addition, Macan (2009) contributed a good overview as well as suggestions on areas for further research on interviews in high-stake selection processes.

The popularity of interviews in personnel selection can be attributed to their generally favourable psychometric properties and high acceptance by both interviewers and applicants. A meta-analytic re-analysis of the relationship of employment interviews and cognitive ability measures by Roth and Huffcutt (2013) found a corrected correlation of .42, suggesting that interviews can capture some variance associated with candidates' cognitive ability. These results support earlier meta-analytic findings by Salgado and Moscoso (2002), who reported a corrected correlation of .41 for general mental ability and conventional interviews. Schmidt et al. (2016) reported a validity of .76 for a combination of cognitive ability tests and structured interviews for the prediction of job performance. It should be noted though that Salgado and Moscoso (2002) also found a lower corrected correlation of .28 between general mental ability and behaviour interviews. Further differences between conventional and behavioural interviews regarding their association with job experience, academic achievement and social skills suggest that different types of interviews could be considered distinct and assess different constructs, thus predicting different aspects of job performance (Salgado & Moscoso, 2002).

Job performance is one of the more widely used criteria in research on the use of diagnostic instruments in personnel selection and development and refers to the organisational value of behaviours people show at work and the results they generate (Hunter & Hunter, 1984; Motowidlo, 2003; Rotundo & Rotman, 2002; Sonnentag, Volmer & Spychala, 2008; Viswesvaran & Ones, 2000). In addition to being linked to the quality and quantity of work outcomes and thus relevant to a company's success, job performance has also been linked to employee well-being and organisational citizenship behaviours (Medina-Garrido, Biedma-Ferrer & Ramos-Rodríguez, 2017; Rotundo, 2002).

There are various ways to measure and predict job performance. These include – but are not limited to – supervisory ratings and 360-degree feedback, capturing organisational citizenship behaviours, work samples, a variety of job-specific indicators based on behaviours or work results (such as sales records) and personality measures (Barrick, Mount & Judge, 2001; Rotundo, 2002; Schmidt & Hunter, 1998; Viswesvaran & Ones, 2000). Meta-analytic research and reviews by Brenner (2019), Hunter and Hunter (1984), Macan (2009) and Schmidt and colleagues (2016) also support interviews as a viable method to predict job performance, with results indicating better prediction of job performance through structured interviews compared to unstructured interviews. One frequently cited result in this context is Schmidt and Hunter's (1998) finding of a corrected correlation of .51 between job performance and structured interviews and .38 for unstructured interviews. Research by DeGroot and Motowidlo (1999) as well as DeGroot and Gooty (2009) has also linked specific visual information in interviews to job performance ratings and interviewer judgements. The association between visual information in interviews and job performance comes with the caveat of possibly being affected by impression management by applicants which is more strongly associated with interview ratings than with job performance (Barrick, Shaffer & DeGrassi, 2009). As such effects are particularly relevant in the context of this work, they will be explored in more detail in later sections of this chapter.

In a more recent review of research on personnel selection methods, Schmidt and colleagues (2016) suggest that interviews measure a combination of cognitive ability, job experience and personality. Based on comparable validity evidence across different studies and application areas, Schmidt and Shaffer (2016) argue that interviews do not vary much from one application to another. In addition to cognitive ability, interviews can also capture candidate personality if the interview process allows for it, or the interview was specifically designed for this purpose (Macan, 2009; Schmidt et al., 2016). Discussing their counterintuitive finding that conventional interviews (which generally tend to focus on candidates' experience, credentials, and self-evaluations) are only moderately associated with academic achievement, Salgado and Moscoso (2002) suggest that interviewers might use questions about academic and work-relevant achievements to gather information about personality or cognitive abilities, especially when interviewing candidates early on in their careers.

One of the key functions of employment interviews is the exchange of information between interviewer and interviewee, wherein both sides attempt to gain a better understanding of each other as well as present themselves (Bangerter, Roulin, König, 2012; Roulin, 2022). Barrick, Swider and Stewart (2010) describe it as "an agenda-driven social exchange between strangers" (p. 1171). As noted by Brenner (2019), it is important to differentiate between interviews as recruitment and as selection tool, as the purpose of an interview can affect the use of different technologies and the psychometric properties of interviewer judgements. In the personnel selection context, Bangerter and colleagues (2012) point out that the information exchange between interviewee and interviewer is competitive as well as cooperative with slight divergence between each side's goals.

Interviews can not only differ in their purpose, but also in their content and processes. One of the key distinctions can be made between conventional interviews focused on a candidate's credentials and self-evaluations, behavioural interviews focusing on their experience and on-the-job behaviours and situational interviews asking candidates how they would behave in hypothetical situations developed from examples of real job-relevant situations (DeGroot & Kluemper, 2007; Latham, 1989; Peeters & Lievens, 2006; Salgado & Moscoso, 2002). In terms of the process, one of the most common differentiations of interview types is based on the degree of structure. Levashina and colleagues (2014) identified 15 components of interview structure of which studies in their meta-analysis on average used six. Despite relatively long-standing consensus that structure improves interviews' psychometric properties, the literature reviewed by Levashina and colleagues (2014) showed a lack of consistency regarding the meaning and definition of interview structure, as well as its operationalisation. Earlier research also focused on the interviewers themselves and whether they had previously received formal interview training and whether this affected the degree of structure in an interview (Chapman & Zweig, 2005). However, the same study by Chapman and Zweig (2005) also found interviewers to be confident in their recruiting and selection efficacy, irrespective of the degree of interview structure.

Self-Other Agreement

The judgement of candidate personality is often one of the goals of interviewers in conventional interviews who may also be using questions related to candidates' professional achievements to indirectly judge cognitive abilities and personality (Salgado & Moscoso, 2002). That there is value in assessing personality in interviews is supported by findings that ratings on the Big Five (c.f., McCrae & Costa, 2008) dimensions agreeableness, conscientiousness and extraversion can explain additional variance in job performance beyond that explained by ratings in situational interviews (DeGroot & Kluemper, 2007). As Powell and Bourdage (2016) point out, the variety of organisational outcomes that have been linked to personality in previous research, coupled with the ubiquity of interviews in personnel selection processes, make accurately judging applicant personality a relevant practical skill for interviewers.

Interpersonal accuracy can be defined as "accurate judgment about any verifiable characteristic of a person or about the group that a person belongs to" (p. 5) through behavior- or appearance-based inferences or recall (Hall, Schmid Mast & West, 2016). One important implication of this, as explained by the authors, is that accuracy itself is a construct and dependent on the operational definition of the criterion and its limitations, such as possible ties to the stimuli. Among the different ways to measure interpersonal accuracy, self-other agreement “reflects a consistency between the internal (i.e., affect, cognition and desire) and external (i.e., behaviour) manifestations of the underlying personality dimension” (Connelly et al., 2021, p. 1356). The correspondence between judges’ perception of targets’ personality characteristics and targets’ actual personality – often measured through self-reports – can be operationalised in different ways, with one of the key decisions being whether the approach to correspondence is variable-centered, i.e., how accurately individual traits are judged across multiple targets, or person-centred, i.e., aiming to correctly judge the personality profile of individual targets (Back & Nestler, 2016; Hall et al., 2018). Which of these two approaches is chosen in a given scenario should be carefully considered, as they appear to represent distinct psychological phenomena (Back & Nestler, 2016; Hall et al., 2018). Self-other agreement is relevant, not only in terms of accurate interpersonal judgements, but also due to its possible intra- and interpersonal consequences, including improved quality of communication and interactions, as well as higher acquaintance and relationship satisfaction (Human & Biesanz, 2011).

Realistic Accuracy Model

The previously referenced Realistic Accuracy Model by Funder (1995) describes the accuracy of a judgement of personality by an observer through the right usage of behavioural cues and is fairly widely used (Letzring et al., 2020). The model specifies four stages that affect judgement accuracy and expands upon prior approaches by including a broad range of elements on both the judges’ (as well as the targets’) side of the process. The steps build onto each other so that if a pervious step is not given, the chain cannot proceed further to – in the words of the Lens Model – generate functional achievement (Funder, 1995; Funder, 2012). The model accounts for this interdependency by specifying a formula with multiplicative relationship for all steps, so that accuracy will be zero if any step is missed.

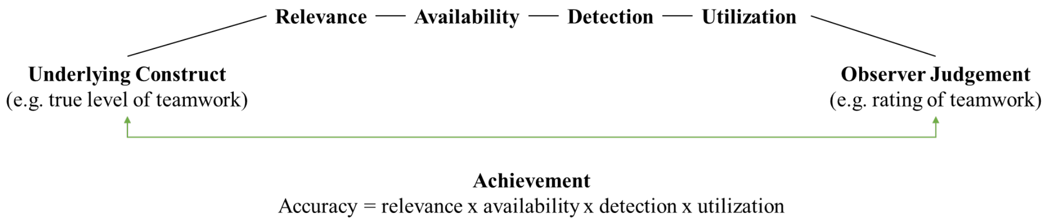
The four steps to generate accurate judgements (functional achievement) are:

- A situation needs to occur in which a participant is showing relevant behaviour that relates to the underlying construct (trait).
- The relevant behaviour needs to be generally available to the observer.
- The observer needs to actively detect the relevant behaviour.
- The observer needs to use the relevant behaviour for judging the underlying construct (trait).

(Funder, 1995)

In terms of the Lens Model (c.f. chapter 3), the Realistic Accuracy Model's *relevance* and *availability* stages correspond to *cue validity*, whereas the detection and utilization stages correspond to *cue utilization* (Back & Nestler, 2016).

Figure 1. Funder's (1995) Realistic Accuracy Model, own adaptation based on Letzring et al. (2020).



In addition to the four steps to generate judgements, the model specifies four moderators of judgement accuracy. These moderators are only briefly outlined here. Letzring et al. (2020) provides a good summarised overview of Funder's (1995; 2012) explanations.

- Good judges who are able to accurately detect and use relevant cues.
- Good targets making relevant cues available to observers through their behaviour. This is related to the concept of judgeability (c.f. Human & Biesanz, 2013).
- Good traits that are associated with a high number of relevant and visible cues. This is related to the concept of evaluativeness (c.f. John & Robins, 1993).
- Good information being available to the judge, both in terms of quantity and quality (for example, through acquaintanceship).

For the purpose of this research, it needs to be highlighted that the Realistic Accuracy Model requires a clear differentiation between observers detecting cues and observers using cues for judgements.

Furthermore, the clear linear formula of this model allows to question the problematic issue of cues being linked to multiple traits and, therefore, to interact on multiple levels. This is an observation that the Realistic Accuracy Model highlights well and it can be discussed further based on the results in Chapter 4.

Self-Other Knowledge Asymmetry Model

The Self-Other Knowledge Asymmetry Model addresses discrepancies between self-ratings and observer ratings in personality judgements. The model postulates that personality traits vary in how observable they are and how much they are affected by unobservable processes, such as thoughts and feelings. As a result, self-ratings are more accurate for traits low in observability and observer ratings are more accurate for traits that are less dependent on intrapersonal processes and presented through behaviours. (Vazire, 2010)

The self-rating and observer rating for one trait are likely to be different, based on differences in the underlying amount of information, as well as the sources of that information which is also referred to as information asymmetry (Vazire & Carlson, 2011). Beyond just information, distortions, such as ego-protective biases, are also postulated to affect self-ratings more than observer ratings; however, this may vary depending on acquaintanceship between observer and target.

The addition of the level of acquaintanceship between observer and target is one of the interesting contributions of the Self-Other Knowledge Asymmetry Model. In line with the model, closer acquaintance improves accuracy for traits with a low observability feature and a higher reliance of knowledge about the intrapersonal processes (Sun & Vazire, 2019).

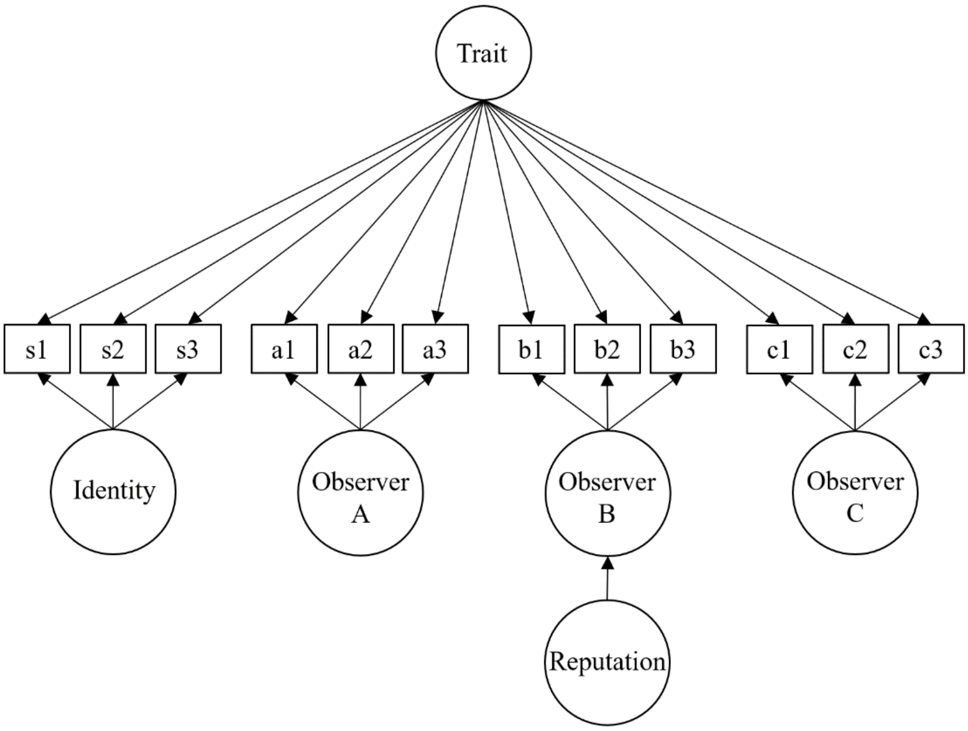
Given these observations, and different to the Realistic Accuracy Model and Lens Model, the Self-Other Knowledge Asymmetry Model allows for full accuracy without the need for overlap between self-rating and observer rating. This is based on the fact that – depending on the observability of a trait and linked characteristics – it can be that only the subject or only the observer are able to access all available information. Hence, full accuracy or a functional achievement could be reached, even though there is no overlap in ratings.

Trait Reputation Identity Model

The Johari Window (Luft & Ingham, 1955) is an interpersonal communication model (Saxena, 2015) that orders the knowledge people have about themselves and others have about them in a two-by-two matrix of possible combinations. The four areas in this matrix (also referred to as regions or quadrants) are: *open*, what's known to others and oneself; the *façade* including what is known to oneself but not others; the *blind* region known to others but not oneself; and what's *unknown* to both the self and others. Based on the Johari Window and the Self-Other Knowledge Asymmetry Model (Vazire, 2010), McAbee and Connelly (2016) developed the Trait Reputation Identity Model, which "posits that a person's standing on an individual trait continuum can be conceived according to the shared versus unique information available across rating sources" (Connelly et al., 2021,

p. 2). These areas in the Johari Window correspond to the elements in the Trait Reputation Identity Model as *Trait* = *Open*, *Reputation* = *Blind* and *Identity* = *Façade*.

Figure 2. Trait Reputation Identity Model, adaptation based on McAbee and Connelly (2016).



One of the strengths of the Trait Reputation Identity Model lies with its ability to model differences between self and other perspectives on personality that may arise due to impression management, context effects in the situation someone is perceived in, stereotypes and the accuracy of communication between different observers (McAbee & Connelly, 2016). The original authors point out that the model can be extended to include multiple timepoints of self-report data or observer data from different contexts or varying degrees of acquaintanceship with the target. In the likely scenario of multidimensionality of personality traits, the model's factors can be supplemented to better reflect the factorial structure of the trait and dimensionality of the items. While potentially problematic, due to confounding the *Reputation* element of the Trait Reputation Identity Model, the model also remains applicable when there is only one rater.

Key Takeaways Regarding Self-Other Agreement

In this section, a few thoughts are highlighted to from the previously-mentioned models, as some of the key points presented in this section are not only relevant to the wider context of this work, but specifically to the subsequent chapters. This includes the possible approaches to the accuracy of personality judgements as focusing either on the accuracy of judgements of individual traits across targets (i.e., variable-centred) or on the accuracy of personality profile judgements across traits within individual targets (i.e., person-centred). Different models on how personality judgements are generated in social interaction situations, such as interviews were presented. Among these, the Realistic Accuracy Model by Funder (1995) is among the most important to the Lens Model-based (Brunswik, 1956) approach of this work, as it postulates that relevant cues need to be available, detected and correctly used to generate accurate judgements of personality. The Lens Model itself will be presented in the subsequent chapter when presenting the specific methods and frameworks used for the studies of this body of work.

When interpreting results on self-other agreement, it is also important to keep in mind that targets and judges differ with regards to the information their respective judgements are based on, as well as the distortions possibly affecting them - as is laid out by the Trait Reputation Identity Model (McAbee & Connelly, 2016).

The Human Factor in Interview Decision Making

Interviewer judgements and, in many cases, corresponding (standardised) ratings are one of the key outcomes of interviews in a selection context. Interview ratings are based on a range of factors, including interviewee performance, which Huffcutt and colleagues (2011) conceptualise as construct mediating between interviewee attributes and interviewer ratings.

Model of Interview Performance

This section outlines one of the more influential models on factors contributing to interview outcomes. The model of interviewee performance by Huffcutt and colleagues (2011) is relevant to the wider context of employment interviews as it has informed research in many of the areas this work draws upon. The model is centred on the construct of interviewee performance which, according to the authors, consists of the content of verbal responses to interview questions, characteristics of how these responses are conveyed and non-verbal behaviours.

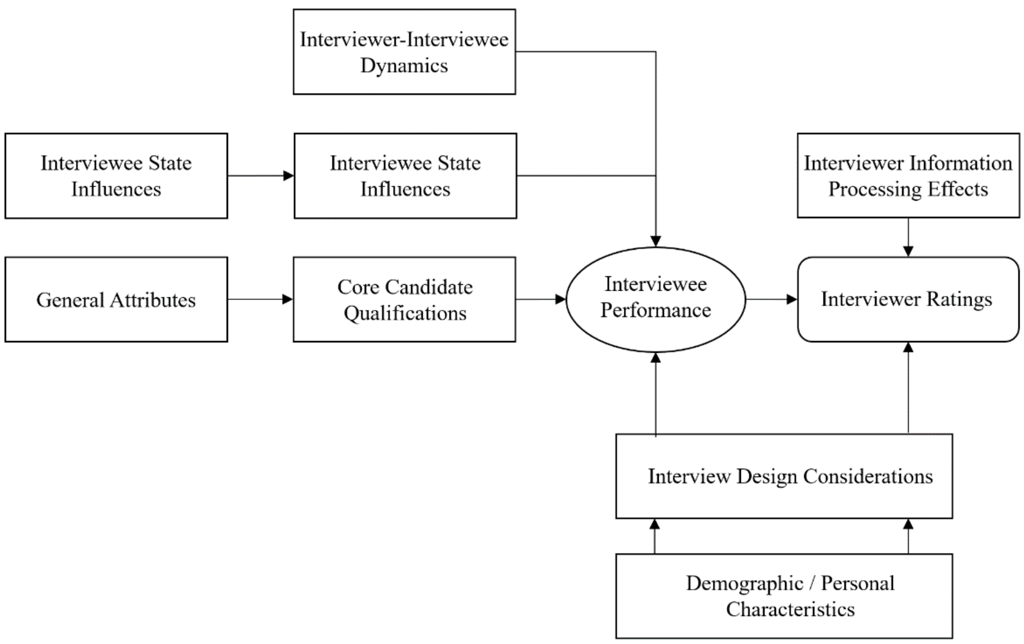
The authors propose a range of factors, including *general attributes*, *core candidate qualifications*, *supplemental preparation*, *interviewee state influences* and *interviewer-interviewee dynamics* that affect interviewee performance positively and negatively in a "give-and-take pattern" (Huffcutt et al., 2011, p. 364). The model does not specify a formulaic representation of these relationships as the authors point out that the same cumulative result can be based on different factor manifestations and combinations.

According to the model, interviewee performance directly affects interviewer ratings. Both interviewee performance and interviewer ratings are proposed to be influenced by the participants' *demographics and personal characteristics* and *interview design considerations*. Interviewer ratings are also subject to influences by *interviewer information processing effects*, which are proposed to contribute to differences between different interviewers' ratings of the same information (Huffcutt et al., 2011).

One of the goals Huffcutt and colleagues (2011) state for their model of interview performance is to promote a shift of focus away from the interview process as key variable influencing interviewer ratings, thus opening up a wide range of possible research avenues on strengthening and inhibiting factors of interviewee performance as part of a more holistic perspective. The model's holistic perspective aligns well with the broad perspective this work adopts on visual information in employment interviews and can be useful in structuring and contextualising its approach and findings.

The model of interviewee performance has informed research in many areas relevant to this work and mentioned in the following sections, such as the effects of cultural differences between interviewer and interviewee, interview structure, impression management, biases, individual differences in social interaction, situational characteristics, as well as differences between types of technology-mediated interviews on interview outcomes (Arseneault & Roulin, 2021; Basch et al., 2021; Derous et al., 2016; Melchers et al., 2012; Swider, Barrick & Harris, 2016).

Figure 3. Model of the Interview Performance (own adaptation based on Huffcutt, 2011).



Influences on the Interviewee Side

Among the various factors influencing interview outcomes particularly relevant to this work are initial impressions or evaluations, which are formed at the beginning of an interpersonal interaction (also referred to as rapport building) (Barrick, Swider & Stewart, 2010; Florea et al., 2018; Swider, Barrick & Harris, 2016). Initial impressions were found to correlate with acceptance decisions (Springbett, 1958) and ratings of interviewee performance, as well as explain additional variance in interview scores beyond expert ratings; however, some research indicates that this effect decreases over the course of a structured interview (Suen, Chen & Lu, 2019; Swider et al., 2016). Physical attractiveness was also found to play a considerable role in the accuracy and positivity of interpersonal judgements at zero acquaintance, leading Lorenzo, Biesanz and Human (2010) to conclude that "overall, people do judge a book by its cover, but a beautiful cover prompts a closer reading" (p. 1777). Frieder, Van Iddekinge and Raymark (2016) found interviewers who engage in rapport building to make their decisions more quickly, suggesting that rapport building might counteract interview structure and caution that decisions made quickly may be inaccurate or unreliable. While a seemingly obvious solution to improve an interview's psychometric properties, limiting or eliminating rapport building (as suggested by Levashina et al. [2014]) may change the social interaction characteristics of the interview and reduce chances for interviewers to gather additional information and to recruit applicants (Chapman & Zweig, 2005; Swider et al., 2016), as well as negatively affect candidate reactions (Brenner, 2019).

One possible underlying mechanism for the effect of initial impressions is the impressional primacy effect, according to which people favour evaluations confirming existing beliefs, such as initial impressions associated with candidate physical appearance in an interview (Suen et al., 2019). Results on how the impressional primacy effect influences interviewer behaviour are mixed, as noted by Dougherty, Turban and Callender (1994), who found first impressions to be related to increased information sharing during the interview and more positive attitudes towards applicants. Contrary to that, Florea and colleagues (2018) found no evidence of confirmatory behaviours related to initial impressions, but their results did show the relationship between initial impressions and interview outcomes to be moderated by the interviewers' need for cognition and accountability. In a similar vein, Springbett (1958) postulates that reviewing application forms prior to an interview triggers an attitude of caution in interviewers, making the more ambiguous information from the interview more likely to be assimilated to the clearer and more defensible information in the application form. Similarly, Torres and Gregory (2018) examined the effect of recruiters reviewing applicants' CVs prior to their responses in an asynchronous video interview and found evaluations to be lower, compared to a condition in which videos were reviewed first. The authors suggest higher expectations created by CVs reviewed prior to videos and lack of background information on candidates based on videos shown prior to CVs respectively to explain this finding. Further adding to the complexity in research on initial impressions is the interaction of primacy and interviewer mood, with positive mood increasing primacy effects due to more assimilative information processing and negative mood reducing them through more accommodative processing (Forgas, 2011). While there is evidence of initial impressions predicting interview outcomes to a certain degree, Frieder and colleagues (2016) caution to keep in mind that interviewee performance may change over the course of an interview and that the content of the interview is also likely to vary in different phases of an interview. Frieder and colleagues (2016) also highlight their finding that decision-making time varied across and within interviewers, with a possible interaction effect between rapport building and the degree of training interviewers received on decision-making time.

Moving beyond initial impressions and towards impression management throughout the interview process, meta-analytic results show notable impact of candidates' impression management tactics on interviewer ratings, with unstructured interviews being more vulnerable to these effects than structured interviews (Barrick et al., 2009). In addition to interview structure, the type of question asked appears to be related to interviewers' ability to detect impression management, with detection being more accurate in situational questions compared to past behaviour questions (Roulin, Bangerter & Levashina, 2015).

The importance of considering impression management in the interview context is emphasised by findings suggesting that impression management is more strongly related to interview ratings than job performance in both unstructured and structured interviews, as this does suggest that impression management may be detrimental to interview validity (Barrick et al., 2009; Roulin et al., 2015). However, it should be noted, that Barrick and colleagues (2009) point out that, for certain jobs, effective impression management could improve job performance. It should also be taken into consideration that candidates can be expected to try and convey a positive image of themselves in employment interviews (Barrick et al., 2009), making it necessary to differentiate between honest impression management (such as emphasising their skills and abilities) and deceptive impression management or faking, which may mislead interviewers' judgements (Bourdage, Roulin & Tarraf, 2018). Taking the possible effects of deceptive impression management into account is important as research suggests it to be more prevalent than interviewers assume (Schneider, Powell & Roulin, 2015) and it can be fairly difficult to accurately detect (Roulin et al., 2015).

In line with definitions of an interview as a social exchange situation (Barrick et al., 2010) and a signalling theory (Connelly, Certo, Ireland & Reutzel, 2011; Spence's (1973) recent findings suggest not only that the amount of impression management used by applicants is predictive of their positive effect and overall performance scores, but also that interviewees and interviewers adapt their impression management behaviour to each other's preferred patterns of responses and impression management (Wilhelmy, Roulin & Wingate, 2021). In the aforementioned study, these preferred patterns of impression management were only related to performance ratings on behaviourally-anchored rating scales by independent raters. In addition to applicant behaviours, interview ratings can also be influenced by applicant characteristics, such as attractiveness, as part of heuristics or biases (Behrend et al., 2012). Suen and colleagues' (2019) findings on the effects of candidate appearance on interview ratings match with Barrick and colleagues' (2009) meta-analytic results that candidate appearance is more strongly related to interview ratings than impression management.

Similar to the debate on whether impression management is a skill relevant to certain jobs or a distorting factor that should be controlled for, Ruben, Hall and Schmid Mast (2015) found evidence of mixed effects of an applicant smiling on interview outcomes, as the type of job may affect the behaviours expected from candidates during interviews. Levine and Feldman (2002) also report differences between women and men in how the eye contact and body posture affect ratings of likability provided by same-sex raters. Specifically in the context of asynchronous video interviews, Basch and colleagues (2021) investigated the effect of preparation time on interview outcomes and impression management and found no increase of performance through deceptive impression management but increased honest impression management. Adding to the body of research showing potentially mixed effects of certain candidate behaviours on interview outcomes is that distracted-looking candidates may be perceived as more extraverted, whereas only maintaining eye contact may result in candidates being perceived as more neurotic (Vijay et al., 2021). One interesting methodological side note to Vijay and colleagues' (2021) study highlighting the possibilities of modern video- and artificial intelligence-based technologies is that it used deepfake videos to simulate and systematically manipulate individual or specific combinations of candidate behaviours, including eye contact, nodding and smiling as part of their stimulus material.

An important research paradigm in the context of non-verbal behaviour and linked to initial impressions with practical relevance for employment interviews are thin slices, referring to very brief, e.g., between 2 and 10 seconds long excerpts of oftentimes recorded interactions (Ambady & Rosenthal, 1993). Thin slices have been shown to be effective for prediction of target personality and other outcomes, such as hireability ratings, with the benefits of additional data in the form of longer or additional recordings quickly diminishing (Borkenau et al., 2004; Murphy et al., 2019). However, Nguyen and Gatica-Perez (2015) found non-verbal behaviour cues extracted from a full interaction to slightly outperform those extracted from thin slices. A more detailed overview of how visual cues relate to the Lens Model by Brunswik (1956) and this work can be found in the corresponding sections in Chapter 3.

Influences on the Interviewer Side

While it is difficult to provide a concise summary of how different variables affect different outcomes generated from different types of interviews due to the sheer number of possible variations, factors and combinations to be accounted for, and wealth of research over decades, one common denominator lies with the interviewers and their task to provide accurate, fair and legally-defensible assessments of applicants.

In the context of impression management, Roulin and colleagues (2015) found interviewers' perceptions of impression management use by applicants to be predictive of their ratings, despite actual impression management-use being controlled for and concluded that "what interviewers see may matter more than what applicants actually do" (p. 433). This statement is relevant in more than one way, as research has shown a range of factors that are frequently subject to biases and may end up disadvantaging certain groups to have potential effects on interview outcomes. These factors include but are not limited to an applicant's gender (Latu, Mast & Steward, 2015; Stamarski & Son Hing, 2015), weight (Agerström & Rooth, 2011; Pingitore et al., 1994; Rudolph et al., 2009), ethnicity (Bartoski et al., 2018; Purkiss et al., 2006; Wolgast, Björklund & Bäckström, 2018; Quillian et al., 2017) and disability (Levashina et al., 2014; Spirito Dalgin & Bellini, 2008; Tagalakis, Amsel & Fichten, 1988), alongside biases (such as the halo effect [Thorndike, 1920]), candidate behaviours (such as impression management) and procedural characteristics of the interview (such as its degree of structure and the possible use of technology). West and Kenny's (2011) Truth and Bias Model of Judgement provides an overview and framework for different influencing factors and how they relate to interpersonal judgements.

In addition to potentially opening up themselves to legal liability through discrimination against certain groups (c.f., Civil Rights Act, 1964; Council Directive 2000/78/EC, 2000), problems in personnel selection processes increase the risk of false hires. This can result in significant costs to companies and negative effects on current employees, as well as the new hire. While there is relatively little empirical research on the costs and effects of false hires (Sutherland & Wocke, 2011), estimates such as the cost of new hires, which the U.S. Small Business

Administration (Weltman, 2022) puts between 1.25 and 1.4 times their salary, can help provide a starting point to assessing the financial impact of wrong decisions in selection processes. Sutherland and Wocke (2011) propose three aspects to be part of the consequences of selection errors: *attribution of the error*; *cost of error*; and *remedial actions*. Research on the cost of error includes estimates between 30% of a hire's first year earning potential and up to five times the false hire's salary in addition to effects such as reduced team morale, client and management goals not being met, loss of clients, weakened employer brand, an increased risk of lawsuits, union activity, litigation fees and accidents which can further increase costs (DeLeon, 2015; Laurano, 2015; Sutherland & Wocke, 2011). In an investigation of turnover-related costs in the healthcare sector, Waldman, Kelly, Arora and Smith (2010) found cost associated with reduced productivity to make up between 1.4 and 3.8% of an annual 500 million USD operating budget and turnover costs to represent between 3.4 and 5.8% of the same budget. The authors also pointed out the differences in turnover costs between different groups of employees, with physicians being the second largest contributor to turnover costs, despite having much lower turnover than all other groups in the analysis due to the high costs associated with individual hires. While determining the average costs associated with a false hire can be difficult due to differences between regions, industries, jobs and individual companies' and employees' circumstances, the above data highlights why continuous research to gain a better understanding of employment interviews is relevant for practitioners just as much as it is for researchers.

Differences in judgement accuracy between raters have been subject to much research and constitute one of the four moderators of interpersonal accuracy proposed in Funder's (1995, 2012) Realistic Accuracy Model. The relevance of considering raters' interpersonal accuracy in the context of employment interviews is highlighted by findings, such as observer ratings of personality being better predictors of job performance than self-ratings (Levashina et al., 2014). A meta-analysis of 103 studies by Schlegel, Boone and Hall (2017) examined the structure of interindividual differences in interpersonal accuracy. Results of this meta-analysis supported a hierarchical model consisting of a higher-order global interpersonal accuracy skill connecting various channel- and domain-specific skills.

An interesting implication of this, as stated by the authors, is that low internal consistency in tests of interpersonal accuracy could actually increase their conceptual strength as they would capture a wider range of facets through moderately related stimuli. A different study comparing the interpersonal accuracy of students and recruiters found that, when looking at target personality profiles, recruiters outperformed students; however, this appeared unrelated to their accuracy when judging individual traits which was interpreted to align with recruiters generally attempting to gain a holistic impression of applicants (Schmid Mast, Bulliard & Aerni, 2011). The finding that interpersonal accuracy may consist of a range of skills aligns with difficulties in attempts to find general characteristics of good and bad judges of personality due to inconsistencies in accuracy across different traits (Allik et al., 2010; Allik, de Vries & Realo, 2016). However, a recent meta-analysis by De Kock, Lievens and Born (2020) identified intelligence and social abilities to be among the more consistent factors influencing personality judges' accuracy. It appears that, not only rater characteristics but also their behaviour in social interactions, particularly the expression of social skills through behaviours such as eye contact or expressions of warmth for example, can affect the accuracy of interpersonal judgements, potentially through facilitating the elicitation of relevant cues from targets (Letzring, 2008). Letzring (2008) also found the quality of judges to affect the accuracy of judgements provided by observers of target-judge interactions. An individual characteristic relevant to both targets and raters is psychological adjustment, as highlighted by Human and Biesanz's (2011a, 2011b) finding that well-adjusted raters tended to judge others to be more similar to themselves compared to less-adjusted raters, while also showing higher normative rating accuracy. There also appears to be a knowledge component to interpersonal accuracy as demonstrated by the effect of knowledge on average personalities on the normative accuracy of individual personality judgements (Rogers & Biesanz, 2015).

Another of the factors influencing interpersonal accuracy is what Funder (1995) referred to as 'good target', describing people who are easier to judge accurately than others (Allport, 1937; Colvin, 1993; Funder, 1995). Target judgeability is important to consider as its variability can affect the judges' ability to form accurate interpersonal judgements (Human & Biesanz, 2013). Colvin (1993) suggests individual judgeability to be influenced by self-knowledge, social skills and psychological adjustment. A review by Human and Biesanz (2013), which also provides an overview of a range of other characteristics, influencing judgeability, reaffirms psychological adjustment as “one of the most consistent predictors of judgeability” (p. 252). While consistency across situations, also described as personality coherence or congruence, as a possible consequence of psychological adjustment (Colvin, 1993; Human & Biesanz, 2013; Sherman, Nave & Funder, 2012) appears to be desirable in terms of judgeability (Human & Biesanz, 2013), Bem and Allen (1974) argue that high variability across situations may be indicative of an individual’s ability to respond appropriately to different situations. In such cases, the authors argue, the individual may be more predictable based on situations rather than traits. However, psychological adjustment appears promote people behaving in line with their individual personality profiles which makes them easier to judge accurately (Human et al., 2014). In line with research and theories on situational strength (c.f. Judge & Zapata, 2015), Sherman and colleagues (2012) found that both individual and situational variables predicted individuals’ degree of congruence between personality and behaviour. Human and Biesanz (2013) note that culture could have an effect on the antecedents and consequences of judgeability through differences in emphasis on independence versus interdependence affecting people’s consistency across roles and situations, as well as differences in cue relevance. Specifically looking at self-other agreement in first impressions, Human and Biesanz (2011a) found psychological adjustment to influence interpersonal accuracy, primarily through increased trait observability and only to a lesser degree through well-adjusted targets’ better self-knowledge. As mentioned earlier in the context of initial impressions in interviews, target physical attractiveness can influence the accuracy of interpersonal judgements in favour of more attractive people (Lorenzo et al., 2010). Possible reasons suggested by Lorenzo and colleagues (2010) include raters being more motivated to understand attractive targets, thus paying more attention to them, and attractive targets providing better information for raters.

The information provided by targets and used by raters to generate judgements was also proposed by Funder (1995) to be one of the moderators of interpersonal accuracy. When examining the difference between factual information and information on targets' values, Beer and Brooks (2011) found neither to be generally superior for personality judgements, though there were slight differences between different traits and both targets and judges perceived value-related information to be of higher utility. The amount of information available for interpersonal judgements also increases the better a target and judge are acquainted (Letzring, Wells & Funder, 2006; Paunonen, 1989). However, increased acquaintance might not only be positive, as raters' personal investment may, in cases of very close acquaintance, distort judgements (Connolly, Kavanagh & Viswesvaran, 2007). Confirming that not only information quantity but also its quality is relevant to the accuracy of interpersonal judgements, Letzring and colleagues (2006) report higher accuracy, consensus and self-other agreement in situations with higher quality interactions.

In addition to differences between judges and targets, as well as information quality, research has investigated the role of trait characteristics, such as visibility and judgeability, and their role in influencing interpersonal accuracy; however, findings tend to be mixed (Connolly et al., 2007; Paunonen, 1989). There is evidence that traits which are more easily observed, such as extraversion, are associated with higher self-other agreement (Human & Biesanz, 2011b). The normativeness of traits appears to have an effect on the relationship between judgement accuracy and assumed judge-target similarity with high and low normative traits being judged less accurately but with higher assumed similarity (Human & Biesanz, 2011b). On the other hand, moderately normative traits tend to be judged more accurately but with less assumed target-judge similarity in an overall pattern of results that could indicate underlying motivational influences to validate one's own highly desirable or undesirable traits or heuristics of raters referencing their self-knowledge more when judging very normative and non-normative traits (Human & Biesanz, 2011b). Relevant in the context of this work is Breil and colleagues (2021) definition of good traits as those that "have, relative to other traits, a high number of valid cues that are also utilized by observers" (p. 27).

Adding to the understanding of methodological factors affecting personality judgements, Allik and colleagues (2010) found that correcting for differences in standard deviation of trait scales led to comparable levels of interrater agreement across different traits, suggesting that trait visibility or judgeability may not be the primary moderator of interrater agreement. This notion is supported by Paunonen's (1989) and Vazire's (2010) findings of an interaction between trait observability and acquaintanceship between judge and target, such that the influence of trait observability on self-other agreement diminishes with increasing acquaintanceship. Results by Biesanz, West and Millevoi (2007) indicate a positive effect of length of acquaintanceship on differential accuracy, although raw profile correlations appeared unaffected and stereotype accuracy decreased. The authors point out that these findings align with predictions in the Weighted Averages Model by Kenny (1991), which is also supported by Letzring and colleagues' (2006) results. A study by Colvin and Funder (1991) also suggests differences in the effects of acquaintance based on the criterion, as acquaintances tended to be better at judging targets' personality than strangers but did not differ in accuracy when predicting target behaviour.

Key Takeaways Regarding the Human Factor in Interviews

Even if technology mediated and without direct interaction (as is the case for asynchronous video interviews), selection interviews and their outcomes are subject to a variety of fundamentally human influences. Among these are interviewees' desire to present themselves favourably through honest or sometimes deceptive impression management, as well as interviewers' being subject to biases, differences in interpersonal accuracy and cognitive processes, such as initial impressions of candidates affecting subsequent evaluations. The four moderators of accurate judgements – good trait; good information; good target; and good judge – proposed by Funder (1995) as part of the Realistic Accuracy Model provide a framework to classify possible influencing factors. The model of Interview Performance by Huffcutt and colleagues (2011) integrates many of these factors as well as processes characteristics of interviews. The previously presented research indicates that judgements based on initial impressions can be fairly accurate and may be leveraged in research in the form of thin slices (e.g., very brief sections of video recordings). The high stakes for both applicants and interviewers should always be kept in mind in the context of personnel selection, as they may also affect motivations, behaviours and outcomes.

Video Interviews

As established at the beginning of this chapter, employment interviews remain one of the most prominent personnel selection methods and are in continuous development, driven by ongoing research and technological advancements (Brandt, Justenhoven & Schöffel, 2020; Brenner, 2019). Particularly relevant in the context of this work are video interviews, as they have seen a rise in adoption throughout recent years, based on their economic advantages over face-to-face interviews, such as reduced travel costs, reduced time to hire, capacity for increased candidate volumes and increased flexibility in scheduling which are becoming more and more necessary in an increasingly globalised workforce (Dunlop, Holtrop & Wee, 2022; Gonzales et al., 2019; Gorman et al., 2018; Joshi et al., 2020; Lukacik et al., 2022; Mejia & Torres, 2018). Despite requiring the applicant and one or multiple interviewers being available at the same time, synchronous video interviews can still greatly reduce the organisational effort compared to traditional face-to-face interviews for both applicants and interviewers. Applicants do not need to travel to an interview location which saves time, potential cost and increases scheduling flexibility (Hickman et al., 2021; Joshi et al, 2020; Mejia & Torres, 2018) and instead they just need to ensure that the parts of their environment visible in the background of the video interview are appropriate (Mejia & Torres, 2018). Companies and interviewers also benefit from more flexible scheduling and save costs and organisational effort by not requiring travel reimbursements or preparation of their premises to provide a room for the interview and possibly a meet and greet for the applicant while on site (Deakin & Wakefield, 2014; Saarijävi & Bratt, 2021). Asynchronous video interviews further increase flexibility for both sides in the interview process through decoupling the applicants' recording of responses to interview questions and their subsequent review by hiring managers. By necessitating pre-defined interview questions and recording times, as well as pre-defined criteria or scales that can be rated by multiple hiring managers, asynchronous video interviews also facilitate a high degree of interview structure (Aon Assessment Solutions, 2017a; Basch et al., 2021; Dunlop, Holtrop & Wee, 2022; Gorman et al., 2018; Lukacik et al., 2022; Mejia & Torres, 2018).

All these characteristics have further accelerated adoption of different types of video interviews in light of events, such as the global COVID pandemic in the early 2020s, which necessitated at least a temporary transition to remote and thus digitalised procedures across many regions and industries (de Villiers, Farooq & Molinari, 2021; Oliffe et al., 2021; Saarijärvi & Bratt, 2021). The above advantages of video interviews remain highly relevant as job markets around the world are experiencing what is referred to as *the great resignation* (Abbasi, 2022; Linzer & Cohen, 2021; Griffiths & Feldman, 2022; Shukla et al., 2022) with up to 41% of employees in a recent Work Trend Index survey by Microsoft (2021) considering changing employer within a year. This spike in job mobility coupled with increased expectations of working from home or in hybrid set-ups "creates opportunities to hire more diverse talent, but … also … a strategy to ensure you don't miss out. And that strategy should include extreme flexibility" (Spataro 2021, as cited in Microsoft, 2021).

While the economic advantages of technology-mediated interviews are fairly easy to quantify and advertise, there is comparably less data on their psychometric properties with regards to the prediction of job performance. Results by Gorman and colleagues (2018) do, however, suggest associations of overall interview performance in asynchronous video interviews with self-reported job performance comparable to findings on video-based situational judgement tests and structured interviews. The authors suggest the structure inherently required to set up asynchronous video interviews and raters' ability to review interviews multiple times contribute to these findings.

Building onto Gorman and colleagues' (2018) research, Brenner (2019) examined the incremental validity of asynchronous video interviews in a high-stake selection process consisting of multiple assessment instruments; however, this study lacked criterion data on job performance. Brenner (2019) reports asynchronous video interviews to be more strongly correlated with assessment centre outcomes than face-to-face interviews, suggesting possible differences in the assessed constructs or effects of the methods themselves. Additionally, Brenner (2019) found evidence of incremental validity of asynchronous video interviews over cognitive assessments and with results also indicating asynchronous video interviews to be less cognitively saturated than face-to-face interviews. Further differences to face-to-face interviews are suggested by asynchronous video interviews explaining additional variance in behavioural outcomes, which Brenner (2019), calling for further research on this topic, states to indicate possible differences in non-verbal cues used in the respective interview modalities.

As the differentiation between synchronous and asynchronous video interviews is not only relevant for practitioners but especially in the context of this work and its research, a short definition of video interviews will be provided in the following sections. More specifically, a two-factor categorisation that allows to differentiate between variants of technology-mediated interviews with various degrees of interaction. For one of those categories, asynchronous video interviews, advantages and disadvantages are discussed and details are described on the specific example of Aon’s asynchronous video interview tool, vidAssess.

Media Richness and Synchrony

As previously established, technological advancement allows for varying degrees of digital mediation in interviews. One approach to differentiating interview modalities that is especially relevant when comparing different types of technology-mediated interviews is via different medias’ capacities to convey information via multiple channels, immediate feedback and personalisation (Daft & Lengel, 1986). While not accounted for in Daft and Lengel’s (1986) initial research and models, their proposition that “Media of low richness process fewer cues and restrict feedback and are less appropriate for resolving equivocal issues” (p. 560) remains applicable to different variants of today’s technology-mediated interviews and can help guide the choice of communication medium, depending on the subject’s complexity. Likewise, the authors’ differentiation of structural characteristics facilitating the use of rich media through face-to-face contact from those facilitating large volumes of data through means of data collection and organisation can still be relevant in the context of employment interviews.

Kock (2005, 2009) proposes media naturalness, rooted in evolutionary psychology, as an alternative mechanism to explain why people prefer richer media. According to the author, interactions with lower similarity to face-to-face interactions and thus lower naturalness require increased cognitive effort, bring higher ambiguity and reduced physiological arousal (Kock, 2005, 2009). Kock (2005) thus places face-to-face communication “at the center of a one-dimensional scale of naturalness where the distance from the center (either to the left or right) could be seen as a measure of decreased naturalness” (p. 125) and postulates an increased likelihood of communication issues the further away from the centre a medium lies.

In line with media naturalness theory, McColl and Michelotti (2019) report recruiters to mention collocation, communication synchronicity and body language as areas of concern in video interviews when compared with face-to-face interviews. While the authors also discuss possible positive effects of experience with a medium on its effectiveness, they do caution that, at the time of their publication, a lack of experience with the medium and technological limitations of video interviews may result in distorted applicant perceptions and ratings (McColl & Michelotti, 2019). The notion to account for media users’ characteristics is also highlighted by Kock (2009) who, from an evolutionary psychology perspective, points out that what constitutes naturalness in a communication medium largely depends on the individual using it and which senses they rely on to perceive their environment. Especially relevant in times of increasing adoption of digitally-mediated communication facilitated by a global pandemic, Carslon and Smud (1999) propose channel expansion to explain why users developing expertise in communicating with specific media may, over time, become better at conveying information and more richly interpreting what they receive.

Building onto the media richness theory, Dennis and Valacich (1999) propose media synchronicity as describing “the extent to which individuals work together on the same activity at the same time; i.e., have a shared focus” (p. 5) in a three-dimensional model of *Media Characteristics*, *Task Functions* and *Communication Processes*. In contrast to Daft and Lengel (1986) and Kock (2005), media synchronicity does not define face-to-face communication as the richest medium but that the optimal choice of medium depends on the specific requirements of a task and situation and that different stages of a process may benefit from the use of different media (Dennis & Valacich, 1999). Dennis, Fuller and Valacich (2008) further develop the media synchronicity theory and reaffirm that most tasks involving multiple people require both conveyance (sharing information) as well as convergence (gaining a shared understanding) processes and thus benefit from multiple media being used.

Dennis and colleagues (2008) further differentiate the media synchronicity theory from earlier media theories through six key differences including the identification of physical media capabilities that affect people's ability to transmit and process information. Ensuring viability of the theory, despite rapid technological development, the authors emphasise that the media synchronicity theory is not based on certain types of media, but rather the features different media may offer which can evolve over time. Further differences highlighted by Dennis and colleagues (2008) are the theory's focus on communication performance rather than media choice, which goes hand in hand with the inclusion of cognitive aspects of information transmission and processing as opposed to prior theories' emphasis of social interaction-related media characteristics, such as immediacy of feedback (c.f., Daft & Lengel, 1986). Based on an overview table of different medias' capabilities by Dennis and colleagues (2008), Brenner (2019) notes that asynchronous video interviews would be considered lower in synchrony than face-to-face interviews. Brenner (2019) also provides a thorough overview of additional media theories, models and frameworks and how they relate to the video interview context, as well as empirical data supporting the importance of interview structure across formats of technology-mediated interviews.

2-factor categorisation of video interviews

Firstly, the general terminology that is used in this work needs to be established. Different authors use a different lens and, thereby, different wording to describe similar but not exactly the same aspects of video interview. For instance, some refer to web-based interviews (Brandt, Justenhoven & Schöffel, 2020), which implies that a video can be captured but focuses on the technology through which the data has been captured. The same applies to the term 'video-conferencing interviews' (Chapman & Rowe, 2002), which again puts the technology first. Others refer to digital interviews (Chamorro-Premuzic et al., 2016), which would include interview forms without visual information, such as telephone interviews (Bauer et al., 2004; Tourangeau, Steiger & Wilson, 2002).

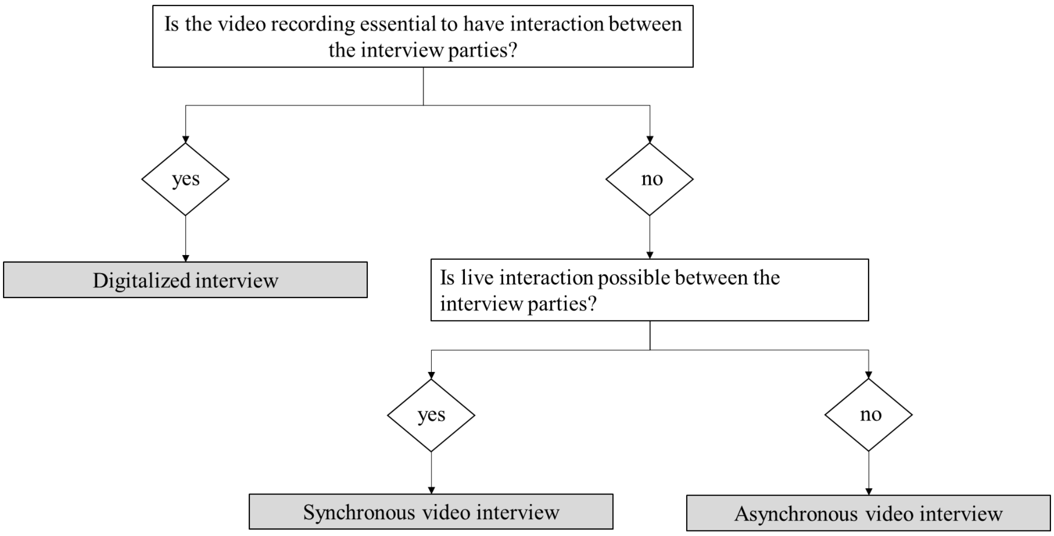
There are two relevant distinctions to be made that lead to the one specific category of interviews that this work focuses on (an overview is illustrated in Figure 4).

The first differentiation is whether or not the video recording is essential for conducting the interview. For instance, traditional selection interviews that use video recording to easier store evidence of the interview are not dependent on the technology to fulfil the main purpose of the interview. The interviewer and observer(s) are able to use other means than the video to conduct their work. This, in turn, provides more freedom to the type of visual cues that can be leveraged during the interview process. Those interviews may be fully or partly digitalised; however, they are not to be seen as video interviews under the definition of this work.

Secondly, the relevant difference that needs to be clarified is between synchronous and asynchronous video interviews, following Levashina and colleagues' (2014) general working definition of interviews in selection processes being an interactive process of two or more interview parties with the specific goal of defining the fitness of the candidate to a job. It is relevant to note that Levashina and colleagues' (2014) definitions of interaction include delayed interaction and this, in turn, includes asynchronous video interviews.

Therefore, following Toldi's (2011) terminology, videos that allow live interaction without delay are referred to as synchronous video interviews while those that are recorded and to be reviewed later are referred to as asynchronous video interviews.

Figure 4. Generic two-factor categorisation of video interview settings.



Asynchronous video interviews

The process flow for asynchronous video interviews is somewhat different to typical video-conferencing settings and follows a more standardised approach (Brenner, 2019; Lukacik, Bourdage & Roulin, 2022). As outlined before, this is an ideal setting to capture, categorise and use visual cues and is the reason why an asynchronous video setting has been chosen. Even though this setting provides specific advantages, there are some limitations – both aspects will be shortly described in the following. However, first, a short definition of asynchronous video interviews.

Pre-recorded interview (Blacksmith & Poeppelman, 2014) and one-way interview (Poh, 2015) are two frequently used synonyms for asynchronous video interviews, as they already capture two of the three major components that differ this setting to synchronous video interviews.

The interviewee is asked to provide a (pre-recorded) video to a stimulus. After completion of the video, it is provided to one or multiple observers to rate the video based on (predefined) scales. This whole process is often fully standardised and – within the application of selection processes – typically administered through a single sign-on platform. Any form of interaction is delayed and therefore not directly related to the interview process itself, but rather the overall selection process (Brenner, 2019).

The following chapter about the tool, vidAssess, will go into further detail to explain a very typical example of such an asynchronous video interview process (Aon Assessment Solutions, 2017a). Therefore, a more generic and abstract description of such a process can be neglected here for the benefit of having a more tangible description on the following pages.

Instead, the most relevant advantages and disadvantages, potentially with regards to visual cues, are briefly outlined. Firstly, the setting allows for the creation of new forms of CVs and video applications (Cammio, 2021) and can help personalise the early stages of a selection process (Torres & Mejia, 2017). This clearly shows that preparation time and allowing interviewees to set themselves into scene can be an advantage for itself to create new selection methods, although it might influence the visual cues that can be observed during a planful pre-recorded video snippet. The freedom for applicants to choose the time and place of recording a video will, in many cases, result in parts of their personal living spaces becoming visible. Lukacik and colleagues (2022) caution that this can give interviewers access to information which may be related to candidates' socioeconomic status and other non-job-relevant data. The authors' conceptual model for the design and use of asynchronous video interviews features such *Applicant Interview Completion Decisions* as directly affecting *evaluation bias*, *evaluator perceptions* and *adverse impact* (Lukacik et al., 2022).

However, reducing involved resources, such as travel time, costs or a person's environmental footprint, are further advantages of asynchronous video interviews (Andrews et al., 2013; Behrend & Thompson, 2013).

Furthermore, the elimination of direct interaction also means no scheduling efforts or costs and each participant can pick a date and time that suits them which can result in reduced hiring times compared to face-to-face interviews (Torres & Mejia, 2017). This is especially relevant if the parties are located in different time zones or have other time-sensitive commitments, such as current employment, when applying for another job (Poh, 2015).

The perceived user experience needs to be addressed as it can also affect people's willingness to apply for certain jobs or accept job offers (Grisworld et al., 2021). Interestingly, there seems to be a decline in both perceived fairness and overall experience whenever the method contains less natural interaction between interviewer and interviewee (Hiemstra et al., 2019). This seems to be the case when comparing live interviews to synchronous video interviews (Chapman, Uggerslev & Webster, 2003), synchronous video interviews to asynchronous video interviews (Langer, König & Krause, 2017) and asynchronous video interviews that are scored automatically and without the help of a human rater (Acikgoz et al., 2020; Gonzalez et al., 2022). These results are further supported by Basch and colleagues (2021) who found the same pattern of differences in perceived fairness between face-to-face and synchronous video interviews to be serially mediated by perceived social presence and impression management. The importance of transparency in this context highlighted by Basch and Melchers' (2019) finding that fairness and usability perceptions of asynchronous video interviews can be improved by providing applicants with explanations that emphasise the benefits of standardisation and flexibility the medium offers.

Research also suggests differences in interview outcomes. Baker, Burns and Reynolds Kueny (2020) found that passive observers in synchronous video interviews rated candidates lower in likability, and hireability compared to a face-to-face interview condition with effects being partially mediated by lower self-reported observer attention. Langer and colleagues (2017), in turn, report an increase of performance scores in asynchronous video interview settings compared to synchronous video interviews. Comparing live and recorded versions of face-to-face interviews and synchronous video interviews, Basch and colleagues (2021) found ratings of the perceived quality of applicant eye contact to mediate the effect of interview condition on performance ratings. The authors interpret this finding to support the importance of non-verbal cues and highlighting the reduced social bandwidth of technology-mediated interviews.

It is often reported for traditional face-to-face interviews that highly structured interviews show higher reliability and validity than interviews with no structure (Schmidt, Oh & Schaffer, 2016). This underlines the findings from Langer and colleagues (2017), given that an asynchronous video interview does need a significant amount of structure in itself. In contrary, Rasipuram and Jayagopi (2016) found that interviewees received lower scores during asynchronous video interviews compared to face-to-face interviews. This is supported by Blacksmith, Willford and Behrend (2016) in a meta-analysis, showing that a higher level of technology usage led to lower scores during interviews. Brenner (2019) links this back to the media richness theory and the difference of bandwidth and synchronicity of (live) face-to-face interviews and asynchronous video interviews. In the context of personality judgements, Wall and colleagues (2013) found that richer media did not necessarily increase judgement accuracy, but rather that the traits to be assessed and the medium used should be matched for optimal fit while accounting for differences in trait visibility across different media.

The last aspect to be highlighted for this setting is the difference in ratings to synchronous options. It becomes clear that multiple observers are able to review the same video material to form their judgement independent of each other and this, in turn, can further reduce any individual observer bias (Eisenkraft, 2013).

Likewise, any bias that would be introduced based on interviewer behaviour is removed (Liden, Marty & Parsons, 1993). Mirroring or impression management is reduced to a minimum and, with the detachment of interviewer and interviewee, the setting creates a more objective, fair and consistent process for all interviewees (Chapman & Rowe, 2001; Liden, Marty & Parsons, 1993).

The variety of opposite findings is too great to be able to adapt learnings from face-to-face interview fully to asynchronous video interviews. However, for itself, this seems to indicate advantages for asynchronous video interviews with clearly shown disadvantages being the low interviewee acceptance.

Automatic Evaluations

Digitally-mediated interviews can allow for automated data collection and analysis; for example, through recordings, automatic speech transcription and capture of data on vocal features or non-verbal behaviours (Hickman et al., 2021; Langer, König & Papathanasiou, 2019). Depending on the type of data collected, analyses carried out and outputs generated, the degree of automation can range from basic ranking of applicants for further evaluation by hiring managers to the adaptive selection of follow-up questions or automatic decisions on who to move forward in the selection process (Chamorro-Premuzic et al., 2017; Langer et al., 2019). An example of a post-hire application of automated data analysis is shown by Speer (2018) who generated narrative sentiment scores from comments in employees' performance reviews and found not only convergence with other performance ratings but also additional explained variance in a range of future performance outcomes.

While the adoption of automatically-evaluated video interviews is increasing, relatively little is known about their psychometric properties, with some evidence regarding their reliability and validity (Hickman et al., 2021). Using data from a simulation, Chamorro-Premuzic and colleagues (2017) point out that the benefits of being able to assess considerably larger candidate pools compared to traditional methods may outweigh the potentially lower predictive validity of automated data analyses. In a study focused on automatic assessment of the Big Five personality traits (c.f. Costa & McCrae, 2008), Hickman and colleagues (2021) found stronger validity evidence for automated video interviews trained using observer reports when compared to those trained on self-reports of personality. Based on these findings, the authors call for more research on the effects of different approaches to modelling decisions through algorithms.

Despite companies emphasising advantages in objectivity and fairness, the methods and data used in the context of automatic data processing should be carefully considered, as they are crucial for the procedure's psychometric properties and the quality of outcomes (Gonzalez et al., 2019; Köchling et al, 2021). Campion and colleagues (2016) provide an overview of steps to create a predictive algorithm based on natural language processing (NLP) intended to emulate a human rater that is, in parts, similar to the approach Aon used for the automatic evaluation component of its asynchronous video interview tool, vidAssess (Aon Assessment Solutions, 2017a). When developing automatic analyses tools, the data an algorithm is based on deserves particular attention, as for example the underrepresentation of certain groups or characteristics in training data can lead to unpredictable effects on resulting classifications, whereas systematic biases in training data tend to be amplified by algorithms applying these patterns to new data (Köchling, et al., 2021; König & Langer, 2022).

While much attention can be given to training data alone, it is important to maintain a holistic perspective, since "bias and unfairness emerge as a result of human decisions made throughout the model development process" (Booth et al., 2021, p. 1), which also includes decisions on the choice and design of an algorithm appropriate for a given use case (Gonzales et al, 2019). In many cases, however, human decisions are part of the algorithm training process based on manually-curated and annotated datasets referred to as supervised learning (van den Broek, Sergeeva & Huysman, 2021). Supervised learning, as opposed to unsupervised learning in which the algorithm independently identifies clusters and connections in data, is the most common approach in organisational high-stakes decisions (van den Broek, Sergeeva & Huysman, 2021). One of the reasons for this is that an unsupervised learning approach can result in the system turning into a black box whose predictions might be accurate but extremely difficult to explain or retrace which can severely limit the explainability of algorithmic analyses (Adadi & Berrada, 2018; Gonzales et al., 2019). Similar to the often-discussed science-practitioner-gap in organisational psychology (Dattner et al., 2019; Gonzalez et al., 2019; Kanning, 2018), Booth and colleagues (2021) point out a discrepancy in the attention paid to bias and fairness issues in psychometric and machine-learning research respectively, with the latter only recently starting to give these topics more consideration. Arrieta and colleagues (2020) provide a thorough overview and taxonomy of explainable artificial intelligence which aims for transparent algorithms and models facilitating human trust and understanding, while considering underlying psychological theories.

On the application side of the process, bias and fairness also need to be considered, especially in terms of how they affect applicant perceptions and behaviour. As outlined by Mirowska and Mesnet (2022), considering the areas of procedural, informational, distributive and interpersonal justice in the context of video interviews and algorithm-based evaluations can be beneficial to disentangle different elements of the process and their effects on justice perceptions. Generally, research indicates that processes relying on automated, algorithmic decisions tend to be perceived less favourably than those relying on human decision-makers (Acikgoz et al., 2020; Mirowska & Mesnet, 2022), especially regarding interpersonal elements of the process such as communication (Gonzales et al., 2022) and perceived fairness (Langer et al., 2019). An interesting nuance in this is that Gonzales et al. (2022) found algorithm-augmented processes that ultimately still rely on human decision-makers to be perceived more positively than solely algorithm-based processes, though still less positive than a human only-based approach. Additional nuance is added by Kanning, Kraul and Litz (2019), who report that their sample perceived companies using algorithmic data analyses in personnel selection as more modern, which contributed to a positive employer image, while simultaneously perceiving them as less attractive as potential employers. Overall, the aforementioned findings tend to align with Mirowska and Mesnet's (2022) results showing a "desire for human elements in the AIE [Artificial Intelligence Evaluation] process" (p. 375), including wishes for a human-like physical representation of the algorithm. According to Gonzales and colleagues (2022), candidate reactions to varying degrees of algorithmic decision-making appear to also depend on candidates' familiarity with the technology, the context of the decision and the criterion used to evaluate candidate reactions. Similarly, van Esch, Black and Arli (2021) suggest that organisations may be able to advertise their use of artificial intelligence to make themselves more attractive to potential applicants, as long as they have generally positive attitudes towards this technology.

In addition to candidate reactions, user reactions to artificial intelligence in personnel selection contexts should also be considered, as reactions to human-generated suggestions can be quite different to those coming from an algorithm, possibly even leading to what has been termed 'algorithm avoidance' despite more accurate predictions compared to humans (Dietvorst, Simmons & Massey, 2015). Such algorithm avoidance can be influenced by factors, such as users' level of knowledge and understanding of algorithms, their perception of autonomy in decision making and the incentivisation of human decision-makers to consider an algorithm's output which, in turn, is also influenced by the perceived algorithm accuracy as this affects potential accountability concerns (Burton, Stein & Jensen, 2020; Logg, Minson & Moore, 2019). The effects of decision-making autonomy and incentivisation on algorithm avoidance also depends on the compatibility between the human decision-making process and that of an algorithm, as well as the convergence of their respective underlying theories and rationales (Burton et al., 2020). The notion that people's reactions to algorithms may also be context-dependent is supported by Logg and colleagues (2019) finding evidence of what they describe as algorithm appreciation when people are tasked with providing visual estimates or predicting geopolitical or business events, music popularity and romantic attraction, even stating that their participants showed a tendency to prefer algorithmic judgements over human judgements (including their own).

Based on the various advantages and disadvantages, combining human and algorithmic decision-makers seems a promising avenue that is currently being explored in research and practice. This approach is also referred to as augmentation and relies on both parties effectively combining their respective strengths (Raisch & Krakowski, 2021). König and Langer (2022) emphasise that human decision-makers will likely remain crucial to personnel selection processes, even if they involve artificial intelligence, as fully automated decision-making may not be legally defensible in certain cases or regulated by regional policies.

Visual Cues in Video Interviews

As established in previous sections, a lot of the factors influencing interview processes and outcomes are based on (or at least influenced by) visual information. This includes interviewers' perceptions of interviewee personality based on specific behaviours (Borkenau & Liebler, 1995; DeGroot & Gooty, 2009; Frauendorfer & Mast, 2015) and is supported by research indicating the effects of available information channels on interview outcomes (Blackman, 2002). The potential additional variability in visual information in technology-mediated interviews is emphasised by a recruiter's comment (cited by Torres and Mejia [2017]), stating that for candidates, asynchronous video interviews can feel like "(…) sitting on the other side of the phone with a friend. The challenge is a lot of people don't see it as a professional interview, so they are dressed in t-shirts or very casual and just not as presentable as they would be if they walked into a face-to-face interview." (p. 694). This quote touches upon multiple aspects relevant to this work. Firstly, it mentions a specific object – a T-shirt – serving as a visual cue to the recruiter.

Secondly, the recruiter mentioned a general impression they derived from the candidate's appearance, i.e., a combination of multiple specific visual cues. Thirdly, an implied standard or expectation that the candidate's appearance failed to meet was expressed. Fourthly and lastly, the quote touches upon differences between face-to-face interviews and asynchronous video interviews with regards to effects on candidate behaviour and appearance, as well as how expectations in these situations might differ between candidates and interviewers.

These elements are also commonly featured in non-scientific sources, such as blog posts, magazines and online articles, that offer advice such as '*7 expert tips for acing your video interview'* (University of Massachusetts Global Administration, n.d.), caution about '*7 common pitfalls of the video-interviewing process'* (TestGorilla, n.d.) or provide instructions on '*Reading candidate body language in a virtual job interview'* (McConnell, 2021). Similar to the recruiter quoted by Torres and Mejia (2017) mentioning a risk of candidates treating a video interview situation too casually, TestGorilla (n.d.) warns recruiters against not ensuring their visible background is neutral and communicating too casually. The latter recommendation is also given to applicants, with particular emphasis being placed on the appropriate background being up to the candidate rather than the company to provide (Borsellino, n.d.; DeCarlo, n.d.; Robert Half, 2021). Addressing the personal appearance aspect, various sources recommend applicants to dress appropriately formal just like they would for face-to-face interviews, while additionally ensuring that the chosen outfit translates well to the limited frame offered by video interviews (e.g., Borsellino, n.d.; Robert Half, 2021). Several of these and similar sources also give recommendations on very specific behaviours and parts of the interview set-up, such as maintaining eye contact by looking at the camera rather than the screen, ensuring an appropriate and natural distance to the camera, maintaining good posture, staying centred on the screen, not excessively moving around but using facial expressions to show engagement, reducing glare through indirect light and ensuring a neutral background (Anderson, 2021; Borsellino, n.d.; Hays, n.d.; Klupacs, n.d.; Meltzer, 2020; National Careers Service, n.d; Robert Half, 2021).

Recommendations for a successful video interview are not always as straightforward. While many sources recommend a neutral background, some suggest that candidates might want to highlight their unique personality through certain decorations or art being visible in the background, thus actively using cues to their personality within their personal living spaces for impression management (Cooper Lomaz, 2020; University of Massachusetts Global Administration, n.d.). Anderson (2021) even suggests sharing a virtual background with all applicants "to minimize so-called "background bias" – dismissive judging of candidates for not having glamorous home settings" (3. Put your candidates at ease section, n.p.).

Taken together, these examples highlight not only the relevance of visual information in the context of video interviews, but also the relevance the type of interview has to both candidates and interviewers and its possible effects on self-presentation and interpretations of each other's behaviours. While mostly high in face validity, the varied and sometimes contradictory research findings and popular recommendations regarding visual data in video interviews, as well as the numerous possible differences to face-to-face interviews (see previous sections), highlight the value of a more systematic understanding of the effects specific visual cues can have on interview outcomes.

Research questions

The present work aims to contribute to research on personality judgements and interviews in a personnel selection context by exploring how visual data in asynchronous video interviews relates to personality judgements. Owing to this exploratory nature, rather than specifying hypotheses, this section will outline some conclusions and expectations based on the research reviewed in the previous sections, as well as questions guiding this work.

One of the overarching themes is the possibilities and limitations of the asynchronous video interview format, as it can both restrict the quantity and quality of available visual data due to the limited field of view (Brandt, Justenhoven & Schöffel, 2020), lack of direct interaction between interviewer and interviewee (Brenner, 2019) and possibly altered applicant behaviour due to lack of familiarity with the situation (Mejia & Torres, 2018), but also give hiring managers access to data not available in on-site interviews, such as glimpses of applicants' personal environments (Lukacik et al., 2022; Mejia & Torres, 2018).

In terms of personality, this work is based Aon's ADEPT-15® model (Aon Assessment Solutions, 2017b) and the Big Five model (Costa & McCrae, 2008) which, in turn, ADEPT-15® is based on. As outlined earlier, the accuracy of interpersonal judgements of personality can be influenced by a variety of factors that are also relevant to the present work. Research by Breil and colleagues (2021) indicates these factors to include the nature of non-verbal cues investigated. The authors point out that interpersonal accuracy was driven more by paralinguistic cues than visual cues. While using the framework of the Lens Model (Brunswik, 1956), a range of cue utilities (visible aspects that video raters are using and are linked to the participant's personality), cue validities (visible aspects that are linked to the self-description of the participant's behaviour) and some functional achievements can be expected for research focusing only on visual cues, validity contributions and correlations are likely to be lower than in multimodal approaches. Further possible limitations, based on Breil and colleagues (2021), include the observed high intercorrelations of cue-utilizations for extraversion, neuroticism, openness and agreeableness which may lead to a number of cues only appearing to be rarely used in a more open, explorative approach partially designed to collect a broad range of possibly available cues within a specific data format and context. However, by aiming to identify and investigate a larger volume of visual cues across five different domains (c.f. Chapter 3), this work follows Breil and colleagues' (2021) call for research examining more comprehensive sets of cues by aiming to collect a broad inventory of cues and methods to expand or reduce it, as well as the call for research on machine-learning-based cue judgements, by focusing on a cue format suitable for automatic detection and rating.

Given the personnel selection context of this work, the majority of interpersonal judgments are likely to be made at zero acquaintance or in the very early stages of acquaintance; although fairly accurate judgements can still be expected, even if based on thin slices of video recordings (Ambady & Rosenthal, 1993; Borkenau et al., 2004; Murphy et al., 2019). As previously noted, trait characteristics can also influence interpersonal judgements; for example, based on trait visibility, observability or evaluativeness (Allik et al., 2010; Funder, 1995; Funder & Colvin, 1988; John & Robins, 1993; Nederstöm & Salmela-Aro, 2014), which affects expectations in this work.

Extraversion has repeatedly been stated to be one of the more observable traits (Borkenau & Liebler, 1992; Funder & Colvin, 1988), with some research even suggesting it is easier to judge from text-only data than neuroticism (Gill & Oberlander, 2003). Confirming this pattern in the context of Lens Model-based (c.f., Brunswik, 1956) investigations of non-verbal cues used for personality judgements, Breil and colleagues (2021) found extraversion to have elicited the highest number of both cues associated with targets' traits and cues used by observers. Examples mentioned by Breil and colleagues (2021) include *cheerful facial expression*, *forward leans*, *gestures*, *relaxed posture*, *attractiveness, neatness*, *lack of eyeglasses* and *lack of dark clothes*, supporting the theory that different categories of cues such as behaviours and appearance are relevant for personality judgements. Nederström and Salmela-Aro (2014) note that the visibility of extraversion-related traits may be increased by the interactive and self-representative nature of employment interviews. Keeping in mind the limited interactivity of asynchronous video interviews but also the chances for self-presentation they provide, e.g., through unrecorded preparation time for individual questions (c.f. Brenner, 2019), similar effects could be expected in this work.

Connolly and colleagues (2007) suggest self-reports as the method best suited to measure neuroticism and agreeableness. Contrary to some previous research but in line with research showing potential for high self-other agreement in judgements on neuroticism-related traits made in assessment centre contexts, Nederström and Salmela-Aro (2014) found high self-other agreement on anxiety-related subscales of neuroticism. Based on the Realistic Accuracy Model (Funder, 1995), the authors propose stress induced by a personnel selection context to maximise variance between applicants. For the present work, these findings emphasise the importance to differentiate between the contexts in which samples were recruited and data was collected, especially considering low versus high stakes situations. Keeping in mind the likely limitations to trait visibility in the context of asynchronous video interviews, it should be noted that Breil and colleagues (2021) did find valid non-verbal cues such as *non-cheerful expressions*, *tense and nervous body language*, *unattractiveness* and *non-neat appearance* being valid and used by raters to judge target neuroticism, supporting the present work's approach. It is worth highlighting, that some of these cues to neuroticism identified by Breil and colleagues (2021) are the inverse of cues found in association with other traits, such as extraversion and agreeableness.

Breil and colleagues (2021) found only non-verbal cues from the categories of body language and appearance to be related to candidates' openness; although raters considered a variety of additional non-valid cues. The present research expands upon the study by Breil and colleagues (2021) by including cues related to environments, which are likely to be considered a source of information by raters (Mejia & Torres, 2018) and to be related to candidate personality (Gosling et al., 2005; Gosling et al., 2002). Gosling and colleagues (2002) mention the distinctiveness of rooms and variety of reading materials as cues with both high cue utilization and validity for judgements of openness.

Breil and colleagues (2021) identified agreeableness to be one of the hardest to judge traits due to fewer non-verbal cues being available to raters with only five cues being both valid and used by observers. It is interesting to note that these cues reported by Breil and colleagues (2021) include *cheerful facial expression*, which the authors also found to be valid and used for judgements of extraversion. Agreeableness being among the harder to judge traits is supported by earlier research reporting self-other agreement for ratings of agreeableness to be the lowest among the Big Five personality traits, especially when target and judge are not acquainted (Ames & Bianchi, 2008; Connolly et al., 2007). Despite a generally low accuracy of agreeableness judgements, Ames and Bianchi (2008) also found agreeableness to be the most judged dimension among the Big Five in the context of initial impressions, particularly when targets were perceived as being higher in hierarchy than the judges. A similar pattern was found by Gosling and colleagues (2002) in the context of personal living spaces with a variety of cues that observers believed to be indicative of occupants' agreeableness but only few that are were actually related to agreeableness measures.

Supporting qualitative data provided by interviewers using video interviews for personnel selection in the hospitality sector, Breil and colleagues (2021) found only cues related to candidate body language and appearance to be valid for judgements of conscientiousness, leading to expectations of similar findings for the present work. The finding that appearance-based conscientiousness judgements are associated with both self-ratings and targets' academic performance further supports this notion (Di Domenico, Quitasol & Fournier, 2013). It is interesting to note that ratings of others' conscientiousness appear to be influenced by a perceived power difference between target and judge, with conscientiousness being more likely to be mentioned when the target is perceived as lower in power or hierarchy (Ames & Bianchi, 2008). Regarding environmental characteristics, Gosling and colleagues (2002) report high utilization as well as validities of cues associated with a room's cleanliness and organisation.

As stated at the beginning of this section, the present work focuses on visual information in asynchronous video interviews and its relation to personality judgements. Informed by previous research, visual information was defined and operationalised as visual cues, referring to specific behaviours or characteristics of candidates and their video interview responses, which can be captured using only visual data. More details on visual cues in the context of this work can be found in

Chapter 3. Personality was captured through judgements made by observers and by targets themselves, as well as their overlap (i.e., self-other agreement).

Both from a scientific and a practitioner's perspective, it is relevant to fully understand the question: what type of visual information is observable in asynchronous video interviews that are recorded for a selection purpose? And how can this visual information be grouped and classified in order to further process and research? The answers to these questions result in the first block of specific questions that are to be tackled with this research initiative. These are:

Q1. Which visual cues can be captured during asynchronous video interviews?

Q1.1. How can the captured visual cues be categorised and classified?

Moreover, given the previously-mentioned link between the method (interview) and the construct (personality) that are typically paired in an employment selection process, it is relevant to investigate the relationship between visual information and personality. More specifically, one can ask the question: how does an observer's judgement of a video interviewee's personality relate to the visual information that is present in the respective asynchronous video interview? For this research initiative, the question is:

Q2. Which visual cues can be leveraged to predict observer ratings of different personality traits in asynchronous video interviews?

Likewise, the relationship between a video interviewee's self-perception of their personality and visual information that they (willingly) reveal is to be put into a descriptive relationship. And we should ask: how do these two dimensions (i.e., self-perception and visual information) relate to each other? These thoughts result in the following research questions:

Q3. Which visual cues can be leveraged to predict self-ratings of different personality traits in asynchronous video interviews?

Q4. Which visual cues can be leveraged to predict both self-ratings and observer ratings for the same personality traits in asynchronous video interviews?

Lastly, and be it just for comparison reasons and opening further investigations specifically for practitioners' use cases, one should ask how the visual information relates to an automatic scoring mechanism of personality. The context of why this research has been triggered and the potential it can set for further research towards more automation and the use of visual information has been given earlier. Therefore, it would be of interest to investigate the relationship between current automatic scoring methods to assess one's personality based on video input and visual information collected on the same video input.

Therefore, the respective research question is:

Q5. Which visual cues can be leveraged to predict automatically-generated personality scores in asynchronous video interviews?

Reviewing the six research questions, the one most prominent and the one to be expected most from is Q4. If this body of research is to help bridge on how visual information can be leveraged in asynchronous video interviews in selection settings to predict personality, the outcome of Q4 is most essential and needed for any steps that are to follow after the completion of this research initiative.

As will become obvious throughout this work, the mindset and direction to create a theoretical framework to work with visual information and its relationship to personality traits are very much in line with Gosling's approach (Gosling et al., 2005; Gosling et al., 2002). This is why the research questions are in line with those that Gosling and colleagues presented in his work published in 2002.

Without wanting to speak prematurely about some of the topics and terminologies, this research initiative – different to Gosling and colleagues' work of 2002 – does use slightly different methods and approaches for some of the steps. Nevertheless, the results displayed by Gosling and colleagues (2002) are to be taken as a blueprint for what can be expected of this work's results. Given that a significant number of visual cues show functional achievement and, with that overlap to explain self-ratings and observer ratings, similar can be expected within this initiative.

Chapter 3: Methods and Instruments

At the start of this endeavour, the work from Gosling and colleagues (Gosling et al., 2002; Gosling et al., 2005; Graham & Gosling, 2011; Graham, Gosling & Travis, 2015; Graham, Sandy & Gosling, 2011) has been paving the way for how this research has been approached. Much of the structure, the research questions and the study design have been leveraged from Gosling's work around the Personal Living Space Cue Inventory by design. The hope is to replicate his approach to how Personal Living Space Cues can be gathered and leveraged to a different setting with slightly different cues (all the while keeping the surrounding attributes of the research).

Therefore, Gosling's key aspects that relate to this research are outlined in the following. Prefixed is the underlying lens model by Brunswik (1956) that Gosling has been leveraged for the Personal Living Space research and which, in turn, will be relevant for some of the empirical discussions as well. The third part of the theoretical framework is the definition of visual cues that is leveraged within this work, accompanied by a visual cue categorisation that is further relevant for the literature review and, therefore, defined upfront.

Following the presentation of the theoretical framework (Lens Model, Personal Living Space and Visual Cues), there is the presentation of two personality models, ADEPT-15 and IPIP, and their respective questionnaires that are leveraged during the empirical studies. Lastly, the topic video interviews with relevant definitions and classifications are presented. These include a detailed description of the tool that is used in the studies for the video recording, vidAssess, and its automatic scoring approach.

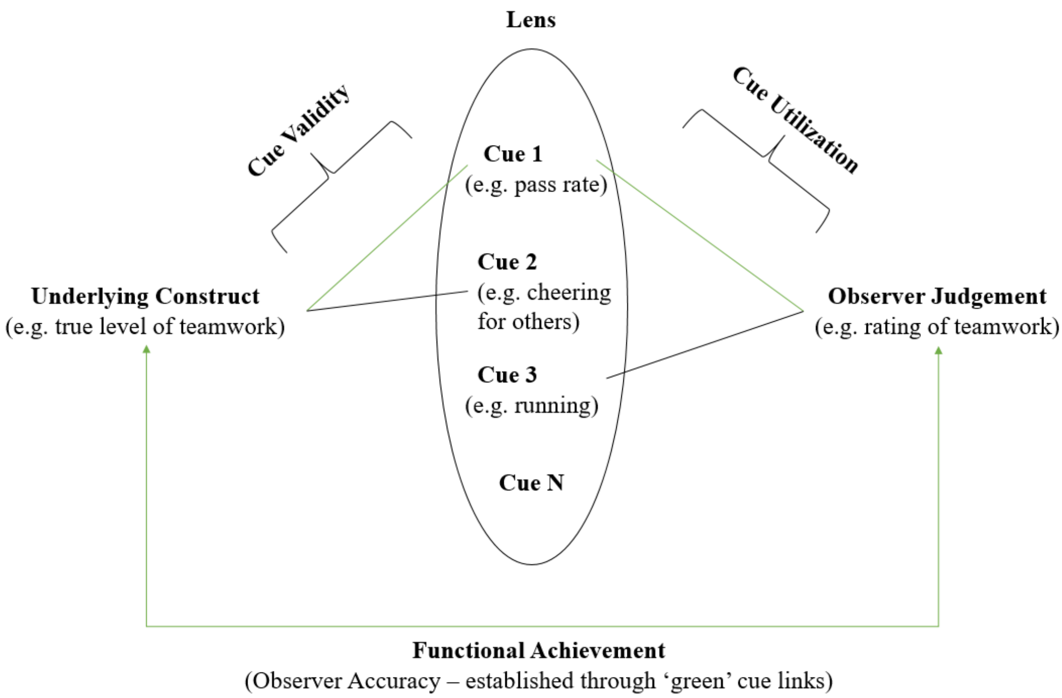
The order of this chapter may seem unusual at first. However, it is set so that knowledge that might be helpful to understand later topics are described early on. For example, the vidAssess section builds upon outlined facts from areas of ADEPT-15 and general descriptions of video interviews. The reader can smoothly flow from one sub-chapter to the next, without the need to go back and forth to understand some of the connections that are drawn or referenced. The accepted downside of this procedure is a tight entanglement between the description of pure theoretical and conceptual models and the description of applied methods and tools.

Lens Model

The Lens Model by Brunswik (1956) is referenced in multiple studies (Borkenau & Liebler, 1995; Breil et al., 2021; Gosling et al., 2002; Hirschmüller et al., 2013) and can act as the conceptual framework to describe how observations, interpretation based on those observations and the correctness of these interpretations interact and relate to each other.

One of the starting axioms of this model is that underlying constructs cannot be seen and, therefore, measured directly. This is much in line with other widely-used latent variable models that have been around for decades (such as those described by Spearman [1904]). According to Brunswik (1956), the observer can use cues as a lens to see the underlying construct. The types of cues are not further specified but can be any element in the environment of the person that is being observed. The observer then forms their judgement of the underlying construct based on the cue information that is available to them.

Figure 5. Brunswik's (1956) Lens Model, own adaptation based on Gosling et al., (2002).



Cue Validity versus Cue Utilization

This work follows the logic from Gosling et al. (2002) that defines the lens to be formed by the sum of different cues that are associated by the observer to form a relationship with the underlying construct. As displayed in Figure 1, the Lens Model shows two links to the lens itself. The link that is established between the lens and the observer judgement is made through the cues that are actually used by the observer to infer the underlying construct. This link is called cue utilization. It does not give any indication about the correctness of the deduction made, rather it describes the observation process.

The other link that is shown is between the underlying construct and the lens. This is referred to as cue validity (Gosling et al., 2002). It defines how well each cue represents the respective level of the underlying construct, which can obviously vary across different cues.

To put this into perspective, the following is a fictional example to illustrate how the lens model works.

Imagine an observer assesses teamwork while watching a player during football practice. In this case, the underlying construct would be teamwork. The player shows different cues during their play, such as to whom they pass the ball or how much they cheer when a teammate scores a point. They also show cues like how much they run to win a ball from an opponent. These are the cues forming the lens which allows the observer to assess the player's teamwork.

In this particular case, the observer uses the pass rates and the amount of running, but not the amount of cheering, which reveals information about the cue utilization for each of the cues. Assuming that the pass rate does link back to the underlying construct teamwork, cue validity is established as well. However, the amount of running is not a cue that reveals anything about the level of teamwork; there is no cue validity for this particular cue.

In this example, the cue pass rate is the only one that shows links to both the underlying construct and the Observer Judgement. With cue validity and cue utilization both being present, the functional Achievement of the model and, with that, Observer Accuracy is established through this cue. However, taking into account the mismatches of the other two cues, the accuracy to assess the underlying construct could have been much greater and, with that, most of the variance could have been ignored if this were an actual case.

Application to this Work

One misalignment that will be interesting to see is to what extent cue utilization is established without the presence of cue validity for those cues. For the present case, it would mean that interviewers and observers use visual cues to inform their ratings about underlying constructs, such as personality traits. However, those ratings would be based on cues that are not linked back to the underlying constructs and thereby carry no validity. The other way around (i.e., cues being present with established Cue Validity link but are not used) is a missed opportunity and both cases can turn into issues when they occur in selection processes.

This model is a great opportunity that allows to bring results, referenced later on in this work, into perspective. The results can be taken and compared to how cue validity and cue utilization and, ultimately, Functional Achievement are present among the cues that are derived within this work and within the video-interviewing setting.

Personal Living Space

Gosling's work on the Personal Living Space and, specifically, the development of the Personal Living Space Cue Inventory has incorporated the lens model (as described in this work). In fact, the Personal Living Space Cue Inventory and, more so, the way it was developed, holds a lot of parallels to this work's approach. Both as a guiding reference model and as a disclosure, the approach that Gosling chose is displayed in this chapter. Highlighted are the aspects that have been taken to be incorporated for this work, as well as those that are deliberately changed within this work to adapt to the differences in setting and cues.

Personal Living Space

No matter what a person does, they leave traces of information about themselves. For example, it could be preferences for certain movies, books or music (Cantador, Fernández-Tobías, I., & Bellogín, 2013; Rentfrow & Gosling, 2006; Gosling et al., 2002) or behaviour such as how someone is talking (Beukeboom, Tanis & Vermeulen, 2013; Mehl, Gosling & Pennebaker, 2006) or what they own (Gosling et al., 2002). Furthermore, the surroundings, the physical and virtual space that an individual creates, displays their own self (Segalin et al., 2017; Gosling et al., 2002). These traces seem to hold valid evidence when it comes to draw conclusions from them in order to infer personality-related information – see Back, Schmukle and Egloff (2008) for examples with email addresses, or Segalin et al. (2017) or, more prominently, Kosinski, Stillwell and Graepel (2013) for their work on social media data.[1]

[1] Looking at the work that Kosinski and colleagues have done with linking one's digital footprint to one's personality in an empirically and data-driven attempt, it could be an interesting endeavour to see if the logic that this work is following and set by Gosling and colleagues to use the lens model,

For the remainder of the work, these traces are referred to as cues, in line with the terminology that has been established in the previous chapter about the lens model. A more detailed discussion on their definitions can be found in the upcoming chapter. However, given that Kosinski's work has already been mentioned, the most relevant specification of cues that are leveraged within this work are visible representations of one's own behaviour – not the (digital) behaviour itself.

These cues can be worked into and displayed in our surroundings in a deliberate way to make conscious statements and form so-called identity claims (Graham, Sandy & Gosling, 2011).

Through identity claims, we can reinforce our own beliefs and create self-directed environments to externally support one's self-perception. They can also be placed so that others perceive them to help transport a specific image about the individual to others and, thereby, help form an impression the individual desires others to have of them. When not deliberately placed, traces, respectively cues, can still appear and are referred to as behavioural residues. These do not necessarily impact the validity of the cues. (Gosling, Gifford & McCunn, 2013)

It becomes apparent that these cues of preferences and behaviours that relate to individual's own self, and, thereby, to their personality, are omnipresent whenever the individual influences their surroundings or outer self. However, they are especially visible when the individual holds sole power of the surrounding, which, in turn, helps to relate back all or most of the visible cues to the individual, whether the cues are displayed deliberately or not. This is given in the concept of a personal living space, which Gosling and colleagues (2005) describe as being "much more than a bedroom but less than a full-fledged house, a PLS (Personal Living Space) is typically a room nestling within a larger residential setting while affording primary territory for a designated individual". (p. 684).

Combining this finding with those from Torres and Gregory (2018), which highlight that video interview recordings most often happen in exactly those rooms,

---

can be used on social media platforms (i.e., the digital personal living space) as well. Once the same framework is established, a comparison between digital personal living spaces and real-life personal living spaces – and more so the rationale for differences – would be an interesting piece of work.

one can draw respective conclusions about the cues that are present during video interview recordings.

On the one side, cues will be present that are deliberately displayed as described by Gosling, Gifford and McCunn (2013). On the other hand, previously-mentioned behavioural residues can also be found in those Personal Living Spaces and linked to, but not exclusively, personality (Graham, Sandy & Gosling, 2011). Through these cues, the person recording the video is trying to make a positive impression and achieve respective positive ratings by recruiters and others who watch the videos (Barrick, et al., 2009). This usage of one's own Personal Living Space and its cues is tapping into typical impression management strategies as is an attempt "to create a particular image in the interviewer's mind during employment interviews" (Roulin, Bangerter & Levashina, 2015, p. 395), as already mentioned in Jansen (2019). Although the typical strategies for impression management do include non-visible aspects such as verbal cues (Roulin, Bangerter & Levashina, 2015), some, if not most, are visible through appearance, behaviour and – in the case of video interviews – environment, i.e. Personal Living Space.

Personal Living Space Cue Inventory

Gosling and colleagues (2005) developed the Personal Living Space Cue Inventory which was "designed to document comprehensively the features of PLSs (Personal Living Spaces)" (p. 684). Furthermore, the aim was "to create an expanded environmental assessment instrument, […] allowing researchers to document the physical features of PLSs comprehensively and effectively" (Gosling et al., 2005, p. 687).

As the development of the Personal Living Space Cue Inventory heavily influenced the development steps of the Visual Cue Inventory of this work, the most prominent steps of the former are described below.

The first phase for Gosling and colleagues was to research established inventories and descriptor tables to collect potential cues. In the next phase, participants of two studies generated new cues that fit their specific setting. In the third phase, the researchers reviewed the combined cue list and established a specific process to narrow the cue list and eliminate cues that did not fit. Lastly, because the researchers realised that the narrowed list of cues was still too long to be used in studies, they proceeded to split the inventory into multiple sections which were to be completed during the studies in a specific order. (Gosling et al., 2005)

Ultimately, the final Personal Living Space Cue Inventory is used by observers (i.e., coders) who can monitor the Personal Living Space and track (i.e., code) what they observed. This cue coding follows a standardised process to ensure it is not dependent on the coder and meets criteria defined by Craik and Feimer (1987) related to reliability and sensitivity (Gosling et al., 2005).

As the attentive reader can observe in Chapter 3, Step 1 is very much in line with what has been described here. There have been, however, a few shortcomings during the empirical studies which are, in turn, picked up in Chapter 4.

However, the cue-coding process itself is more different and this is due to somewhat different settings. Gosling and colleagues used multiple coders for the same Personal Living Space: (1) to ensure reliability through interrater agreements; and (2) to handle the variety and number of cues. The upcoming studies, specifically study 3, will contain only one coder per video, which is the equivalent to Personal Living Spaces in Gosling's settings.

Further modifications from the guidelines of Gosling's work around the Personal Living Space Cue Inventory – for better or worse – that have been executed during this endeavour are picked up and thoroughly discussed in Chapter 4.

Visual Cues

Definition and Differentiation to Similar Constructs

The cues that are used in this work will include (but not solely consist of) those from the Personal Living Space Inventory. Therefore, it needs a specific categorisation and inclusion, as well as exclusion criteria of what type of cues are being collected and, thus, will form the Visual Cue Inventory.

Before going on to the specifics of what and how visual cues are defined within this work, the following abstract is a short detour as to why – of all the cues that could have been focused on during an asynchronous video interview setting – visual cues are focused in this work.

The construct(s) that are focused on to form the lens models are the different personality factors of the personality models that are used and further described in the following two chapters. However, when it comes to the central part of the lens model, the cues, DePaulo (1992) highlights the significance of visual cues given the lack of a person's control and thereby their revealing power of one's true personality. Ambady, Bernieri and Richeson (2000) suggest that visual cues are forming the baseline for observers when rating personality. Both those aspects highlight the potential of visual cues being relevant for both the cue validity and cue utilization. Despite the relationship between visual cues and personality being key to various publications, an overarching definition for visual cues is not (yet) present (Wall & Campbell, 2021).

When setting general inclusion criteria for visual cues for this work – and with that the definition of what is to become a visual cue – it becomes apparent that a multitude of different types of cues are present during an asynchronous video interview. Some have been picked up in recent studies by Cannata and colleagues. For instance, to investigate and form an integrative framework that includes both traditional measurements of personality judgements and personality computing as researched in disciplines of artificial intelligence (Cannata et al., in press).

Within this work, the focus is to remain on purely visual present cues. Ones that do not require any form of audio information. This explicitly excludes prosodic elements or any non-verbal cues that have visible components (e.g., loud gulping).

Furthermore, visual cues that are to be collected will include both those that are visible during an entire video and those that are only present for a short period of time (due to a change of behaviour of the video interviewee or similar change of setting). This generic differentiation is guided by Gosling's categorisation of cues into global descriptors and specific content. Applied to this work, the split will be into static cues (those that are omnipresent during the video take) and dynamic cues (those that require a change during the video).

Following this logic, dynamic cues may vary in their frequency (i.e., how often something happens) while static cues differ may vary in their intensity (i.e., how exposed or how bright something is). Ideally this allows for wide scales and, therefore, variance and interesting data modelling once cues are identified and coded.

Visual Cue Categories

Even though a broad categorisation for visual cues is already made (static versus dynamic), a different form of categorisation is needed – one that is content-driven and more to the nature of the visualisation. This approach is relevant to the author for three major reasons which, in turn, have been influenced by Gosling et al. (2002), Breil et al. (2021) and Mejia and Torres (2018).

Firstly, the visual cue identification phase will be explorative and, to some extent, qualitative. Directing participants to specific aspects of videos or getting an overview of which areas of visual cues might still be underrepresented during the visual cue collection phase (i.e., Step 1 in the upcoming Chapter 3) is likely needed. This, in turn, is easier to do if a categorisation system is in place. Furthermore, it helps with setting up details, such as instructions and prompts for the studies in the first place.

Secondly, the other part of the visual cue collection period is a research of systematic literature. Identifying keywords and being specific in what and where to search for and classifying literature are essential. Therefore, a categorisation system is helpful in efficiently running the literature review.

Lastly, this categorisation system allows an easier and structured follow up when further investigating and building upon the visual cue inventory. Categories can be individually modified to branch off specific research work items. For instance, additions to categories can be made and cues can be newly coded and investigated without disturbing the integrity of the inventory as such.

This body of research has a total of five categories in which visual cues are to be classified.

The first category will include all dynamic cues that relate to the face, which is expected to almost always be fully visible during the video sequence. In addition, the second category will include all other dynamic cues that relate to the video interviewee directly but does not relate to the face. However, different to the face, which is almost always fully visible, the same cannot be said about the remaining body. Great variance can be expected from full-body visibility and, with that, a greater number of visual cues could be picked up to face-shot only where no body parts are visible. Any static cues that directly relate to the video interviewee and, therefore, relates to their general appearance are grouped into the third category.

The last two categories do not directly relate to the video interviewee but are influenced by them. The fourth category will include any visual cues that relate to the video medium as such and any form of media properties that can be observed. Given both the preparation and control about the setting from the video interviewee, as well as the influence of presentation to observers, this category might not be a regular appearance in similar literature and studies. However, it is likely to be one that will become more prominent over time.

The fifth category is somewhat the essence of the Personal Living Space Inventory and will cover any surroundings and environmental (visual) cues. All categories and a short description for each category are displayed in Table 1.

Table 1. Visual cue categories and their description.

#	Category	Description
1	Face	Dynamic cues related to video interviewee’s face.
2	Body	Dynamic cues related to all body parts except the face.
3	Appearance	Static cues related to video interviewee.
4	Media Properties	All cues covering the video media properties.
5	Environment	All cues related to video interviewee’s environment.

ADEPT-15

The ADEPT model and assessment are developed and distributed by Aon Assessment Solutions. Both pieces are tied together and are referenced and used within this work. The next two sub-chapters will dive deeper into explaining the details around the personality model of ADEPT and the assessment that is used to measure the ADEPT model.

The first reason for using the ADEPT model and questionnaire in this work is that it was easily available to the author and can be leveraged free of charges given the affiliation between the author and Aon Assessment Solutions. However, no other Aon representative than the author himself has influenced or took part in how Aon's property has been leveraged so that objectivity can be ensured. All actions have been decided by the author and either executed directly or indirectly through instructions towards the study supervisors as mentioned at the respective steps.

Moreover, the second and more prominent reason for including ADEPT in this work is to allow insightful next steps. The video data is collected through the vidAssess tool, which allows users to generate scores automatically, leveraging a Natural Language Classification (NLC) scoring. The underlying model and the data that the algorithm was trained with is ADEPT based as well. Therefore, any further analyses (and where self-rating and automatic scoring of the video data will be relevant) is easier to be used given the closeness between those two variable types, given that they are supported by the same conceptual framework (i.e., ADEPT).

The Model

Even though there are multiple movements as to how personality can be conceptualised, most of the past 30 years of research align to the Five Factor Model of personality. This divides behaviour into five factors or traits. In this work, trait and factor are used synonymously. This model, often referred to as the Big Five Model, has been widely researched and validated in all different settings, languages, cultural contextualisations and similar variances (Costa & McCrae, 1985; Costa & McCrea, 1991; Costa & McCrea, 1992, Costa & McCrea, 2005).

A well-structured definition of each of the five traits can be found in McCrea & John (1992), which has also been used as the foundation for the ADEPT model definition. The authors of ADEPT, Boyce, Conway and Caputo, further leveraged the approach first outlined by DeYoung, Quality and Peterson (2007) to split each factor into two dimensions each, using this overview as a starting point and baseline model. (Aon Assessment Solutions, 2017)

However, Boyce et al. desired to add a leadership-specific trait to the ADEPT model, having in mind a business-related application to the model for high-level individuals and executive leaders. Following this thought, the lexical approach to develop the Five Factor Model was replicated in smaller scale by proceeding with the following steps:

(1) Collecting of adjectives and general descriptions that Boyce et al. had access to through Aon Assessment Solutions' database, such as assessment ratings or interview reports.
(2) Excluding any phrases that relate to any of the 10 dimensions of the existing baseline Five Factor Model.
(3) Grouping of the remaining phrases based on their similarities and form respective new dimensions.

This procedure produced five additional dimensions, that are not yet covered by the baseline model. Two of which form the sixth factor of the ADEPT model, while the other three dimensions are added to existing factors. This results in ADEPT consisting of six factors or traits, and a total of 15 underlying dimensions, 10 in line with what DeYoung and colleagues reported and five additional dimensions. Moving forward, the ADEPT model (and questionnaire) will also be referred to as ADEPT-15, which is the trademark name to both the model and the questionnaire. (Aon Assessment Solutions, 2017)

Table 2 displays an overview of the six traits and respective dimensions within each trait, as well as a comparison to the baseline model that was used. An overview of definitions for each ADEPT dimension is shown in Table 3.

Table 2. Overview of ADEPT traits, ADEPT dimensions and how they relate to the Five Factor Model as described in Appendix I.

ADEPT traits	ADEPT dimensions	Five Factor Model relation
Adaptation Style	Conceptual	Openness to Experience
	Flexibility	Openness to Experience
	Mastery	Newly Added/Not Related
Task Style	Structure	Conscientiousness
	Drive	Conscientiousness
Interaction Style	Assertiveness	Extraversion
	Liveliness	Extraversion
Teamwork Style	Sensitivity	Agreeableness
	Cooperation	Agreeableness
	Humility	Newly Added/Not Related
Emotional Style	Composure	Emotional Stability
	Positivity	Emotional Stability
	Awareness	Newly Added/Not Related
Achievement Style	Ambition	Newly Added/Not Related
	Power	Newly Added/Not Related

Table 3. ADEPT dimensions and their definitions as described in Appendix I.

ADEPT dimensions (traits)	**Definition**
Conceptual (Adaptation)	This aspect of personality measures the extent to which someone is conceptual and intellectually curious. High scorers tend to be inquisitive and philosophical but may be overly abstract and unrealistic. Low scorers tend to be conventional with less curiosity, as well as more concrete and practical.
Flexibility (Adaptation)	This aspect of personality measures the extent to which someone is flexible, adaptable and open-minded. High scorers tend to be open to new ideas and experiences but may come off as inconsistent or indecisive. Low scorers may be inflexible and set in their ways, but more predictable as they seek tried-and-true approaches.
Mastery (Adaptation)	This aspect of personality measures the extent to which someone is learning-oriented and improvement-focused. High scorers tend to be focused on self-development, practice, and the belief that others can improve; though may be unrealistic in their views of others or their own potential. Low scorers are less concerned with continual self-development, and believe people do not often change much, but they can focus on getting done what is needed.
Structure (Task)	This personality aspect reflects the extent to which someone is planful, detail-oriented, and rule-conscious. High scorers tend to be careful, safe, and orderly, but also perfectionists. Low scorers tend to be disorganized and easily bored, yet can find innovative ways to handle problems and are more likely to focus on the big picture.
Drive (Task)	This personality aspect reflects the extent to which someone is proactive, and persistent. Those who score

	high tend to be reliable, hard working, and accountable, but may get overly focused on narrow goals and can be seen as rigid. Those who score low tend to be reactive and less deadline-oriented, but can shift more easily from goal to goal.
Assertiveness (Interaction)	This aspect reflects the extent to which someone is assertive, decisive, and competitive. High scorers are persuasive and bold, but can be confrontational and aggressive. Low scorers are less concerned with winning and are more cautious when making decisions. Also, they prefer to avoid conflict and may give into others too easily.
Liveliness (Interaction)	This aspect of personality focuses on the extent to which someone is outgoing, energetic, and socially confident. High scores tend to be sociable and friendly, though they may be rambunctious and attention seeking. Low scorers tend to be more reserved and quiet, but also more private and unlikely to offend to others.
Sensitivity (Teamwork)	This personality aspect reflects the extent to which someone is compassionate, caring, and understanding. Those who score high tend to be warmhearted, patient, and tolerant, but may have difficulty providing negative feedback or being firm with others. Those who score low tend to be stoic and tough-minded, but also candid and direct.
Cooperation (Teamwork)	This personality aspect reflects the extent to which someone is cooperative and trusting. People who score high tend to be team oriented and accommodating, but can sometimes be taken advantage of by others. Those who score low tend to be less interested in teamwork, but are also more independent-minded.

Humility (Teamwork)	This aspect of personality measures the extent to which someone is modest and genuine. High scorers tend to be humble and unselfish; but they may be less effective in advocating for their own interests. Low scorers are proud, cunning, shrewd, and can be manipulative; but are also bold and can be proficient at managing situations requiring tact and posturing.
Composure (Emotion)	This personality aspect reflects the extent to which someone is composed, calm, and relaxed. High scorers tend to be tranquil, retrained, and calm under pressure, but can seem aloof and detached. Low scorers tend to be impulsive and excitable; but also demonstrate passion, excitement, and enthusiasm.
Positivity (Emotion)	This personality aspect reflects the extent to which someone is happy, optimistic, and resilient. High scorers tend to be hopeful and positive, but may downplay or disregard potential problems. Low scorers can be pessimistic and overwhelmed with obstacles, but tend to be more pragmatic. Low scorers also are effective advocates for unpopular decisions.
Awareness (Emotion)	This aspect of personality measures the extent to which someone is reflective and self-aware. High scorers are introspective and know their own strengths and weaknesses, but may be self-absorbed. Low scorers have a static self concept and are resistant to feedback, yet are less concerned with or care what others think about them.
Ambition (Achievement)	This aspect of personality measures the extent to which someone is ambitious and goal-directed. High scorers are relentless in their pursuits, but can be obsessive and are rarely satisfied. They may also pursue individual goals in

	lieu of team goals. Low scorers are satisfied with their current status and often have a good work-life balance.
Power (Ambition)	This aspect of personality measures the extent to which someone is controlling, directive, and motivated to lead. High scorers tend to be interested in leadership, control, and influence. Low scorers tend to be team players, lead by example, and willing to let others to take control.

The Questionnaire

The ADEPT-15 questionnaire was developed by Boyce and colleagues to measure the ADEPT model and be "(1) reliable and valid; (2) resistant to faking and impression management; (3) applicable for global use; and (4) secure and efficient in unproctored, high-volume settings" (Aon Assessment Solutions, 2017, p.15). With those guiding principles, the decision was made to use a Computer Adaptive Testing procedure, which allows for minimal test length, high measurement accuracy and – given the context the questionnaire is to be applied in – high test security (Hambleton, Swaminathan, & Rogers, 1991). Participants completing the questionnaire will receive the next items based on previous responses.

In addition, a pairwise, forced-choice format is chosen for displaying the statements to which participants would need to respond, keeping the cognitive effort lower than a comparison between multiple statements (Vasilopoulos et al., 2006) while also reducing the effort of impression management or faking good (Christiansen, Burns & Montgomery, 2005; Converse, Peterson & Griffith, 2009).

The scoring model for the ADEPT-15 questionnaire is based on the Multi-Unidimensional Pairwise Preference scoring model which was developed by Stark and colleagues and reported in Stark, Chernyshenko & Drasgow (2005) and Drasgow, Chernyshenko & Stark (2009).

Figure 3 shows an example item which is constructed of two statement pairs that load onto different ADEPT dimensions. For each item, the participant needs to choose which statement they agree with more. Each questionnaire administration will contain 100 of these items, picked from a pool of 1,500 statements based on the current scoring values per dimension. Once an administration has been completed, the six ADEPT traits are derived from the 15 dimensions per participant.

The questionnaire is not freely available and needs to be requested from Aon Assessment Solutions; however, it is free of charge for research initiatives and educational purposes.

Figure 6. Example item of the ADEPT-15 questionnaire.

IPIP.

The previous chapter highlighted the rational and good reasons why ADEPT-15 is leveraged and used within this work. For different reasons, the author made the deliberate choice to further include another measurement for personality. This chapter will highlight those reasons and shortly explain the set-up and modus operandi of the International Personality Item Pool (IPIP).

Reasons to Include IPIP

Apart from being able to answer the research questions and thereby moving further along with this research initiative, another helpful contribution that this work is hopefully generating is the dataset for the main study (step 3). There are multiple additional perspectives that can be taken, where this dataset can be leveraged – individually as well as part of a meta study. To enable an easier transfer for other research fellows, the IPIP, as a more established personality measurement, is included (Goldberg, 1999). The IPIP is an open-source inventory, therefore freely accessible and will hopefully inspire others to continue using the generated dataset – even if they are unfamiliar with the ADEPT-15 model.

The second reason why the IPIP model is included in this work is its established use for self-rating and zero-acquaintance rating (Goldberg, 1999). As can be read in the previous chapter, the ADEPT-15 questionnaire is a self-rating questionnaire with an answer-dependent, adaptive item display process and respective adaptive scoring. This makes it much more difficult to convert it into an observer rating tool. To date, no official zero-acquaintance rating tool based on the ADEPT-15 model exists. It is possible to directly ask observers to review the ADEPT trait definitions and provide normative, liker scale-based ratings without the further use of items. However, this does not incorporate the full vision of the ADEPT-15 model, which is why this is valid to do; however, it remains a second-class option for academic purposes. Using the IPIP inventory both in self-rating and observer rating form allows for less confounded deductions.

Lastly, the addition of the IPIP model to this work and also to the final dataset allows for further research between the ADEPT-15 model and another classical

Five Factor Model. There have been some (but no published) construct comparisons between the ADEPT-15 model and other Five Factor Model questionnaires. This work will enable a dataset with which further research in this direction will be possible. Some analyses will be part of this work; however, a full in-depth comparison is not in scope. For instance, it might be interesting to gather evidence on the conceptual work carried out by Boyce and colleagues to add another factor and how much additional insights this generates. Likely even more interesting might be follow up work on the additional dimensions that have been added to existing factors and how much unexplained variance they can uncover compared to the apparently narrower traditional factor definitions.

These and similar comparisons have not yet been published. Therefore, the dataset which is to be generated – and, importantly, to be used for future research – creates a unique opportunity for academic research on the ADEPT-15 model, fully independent from influences of the company owning full rights on ADEPT-15.

The Model

The IPIP is an open-source pool of items for personality measurement consisting of 3,320 items in total. These items cover a total of 463 different scales (Goldberg, 1999; Goldberg et al., 2006). On the website (https://ipip.ori.org), the authors describe the intention behind its development as follows: “This IPIP website is intended to provide rapid access to measures of individual differences, all in the public domain, to be developed conjointly among scientists worldwide. Later, the site may include raw data available for reanalysis; in addition, it should serve as a forum for the dissemination of psychometric ideas and research findings." (https://ipip.ori.org/HistoryOfTheIPIP.htm)

Usage of the Mini-IPIP Questionnaire

The Mini-IPIP is a 20-item variant of the IPIP Five Factor Model measure developed by Donnellan, Oswald, Baird, and Lucas (2006). The Mini-IPIP measures the Big Five personality traits with four items each. In a series of studies, the authors established that the Mini-IPIP scales showed good internal consistencies and associations with the Big Five Facets comparable to those of the 50-item IPIP-FFM scales. Re-test reliability is also reported to be similar across the 50- and 20-

item versions. Donnellan and colleagues conclude that, from a practical perspective, the benefits of shorter inventories, namely participant experience and motivation, may, in some cases, outweigh disadvantages with regards to psychometric properties compared to their long-form counterparts.

For the self-rating part of the data collection in step 3, the original and unedited Mini-IPIP items (as provided by Donnellan, Oswald, Baird, and Lucas [2006]) were used. These were set up as a HTML questionnaire on Aon's assessment platform, mapTQ, and presented to participants along with the other used tools, i.e. ADEPT-15 and vidAssess.

As for the observer ratings for the dataset in step 3, the Mini-IPIP was adapted for raters to judge vidAssess participants' personalities. The recommendation for adapting IPIP items for judgements of others is to convert them into the third-person format by adding an ‘s’ to verbs and changing any references to the first person to the third person. (https://ipip.ori.org/Third-Person-Items.htm)

The Mini-IPIP items were changed as described above and reviewed by two Aon internal SMEs. Feedback included items such as ‘Gets chores done right away’ being deemed difficult to rate in the context of vidAssess candidates responding to a work-related interview question. The items to replace them were chosen from the remaining 30 items in the full version of the IPIP. It should be noted that this is likely to have affected the scale's psychometric properties as the selected items may target slightly different areas within the facets than those selected by Donnellan and colleagues.

In addition, the number of reverse-coded items was changed for all facets. The number of reverse-coded items in the original Mini-IPIP was two for each scale, except for Intellect which includes three. The altered version of the Mini-IPIP included no reverse-coded items for Agreeableness and Intellect, one each for Conscientiousness and Extraversion and three for Neuroticism.

Additional feedback was provided on the phrasing, where the SMEs stated items that included a reference to the person (e.g., ‘They pay attention to details’ instead of ‘Pays attention to details’) were more intuitive to understand and rate.

The final list of items and respective details can be found in Table 4.

Table 4. Details of the adjusted and used Mini-IPIP questionnaire.

B5 Dimension	Item	Reverse Scored	Source
A	They sympathisze with other's feelings		Mini
A	They are not interested in other people's problems	R	Mini
A	They feel other's emotions		Mini
A	They are not really interested in others	R	Mini
C	They pay attention to details		Full
C	They do things in a half-way manner	R	Full
C	They like order		Mini
C	They make plans and stick to them		Full
E	They feel comfortable around people		Full
E	They don't talk a lot	R	Mini
E	They find it difficult to approach others	R	Full
E	They keep in the background	R	Mini
I	They spend time reflecting on things		Full
I	They are not interested in abstract ideas	R	Mini
I	They have difficulty understanding abstract ideas	R	Mini
I	They love to think up new ways to do things		Full
N	They are not easily bothered by things	R	Full
N	They are relaxed most of the time		Mini
N	They get stressed out easily	R	Full
N	They rarely get irritated	R	Full

VidAssess

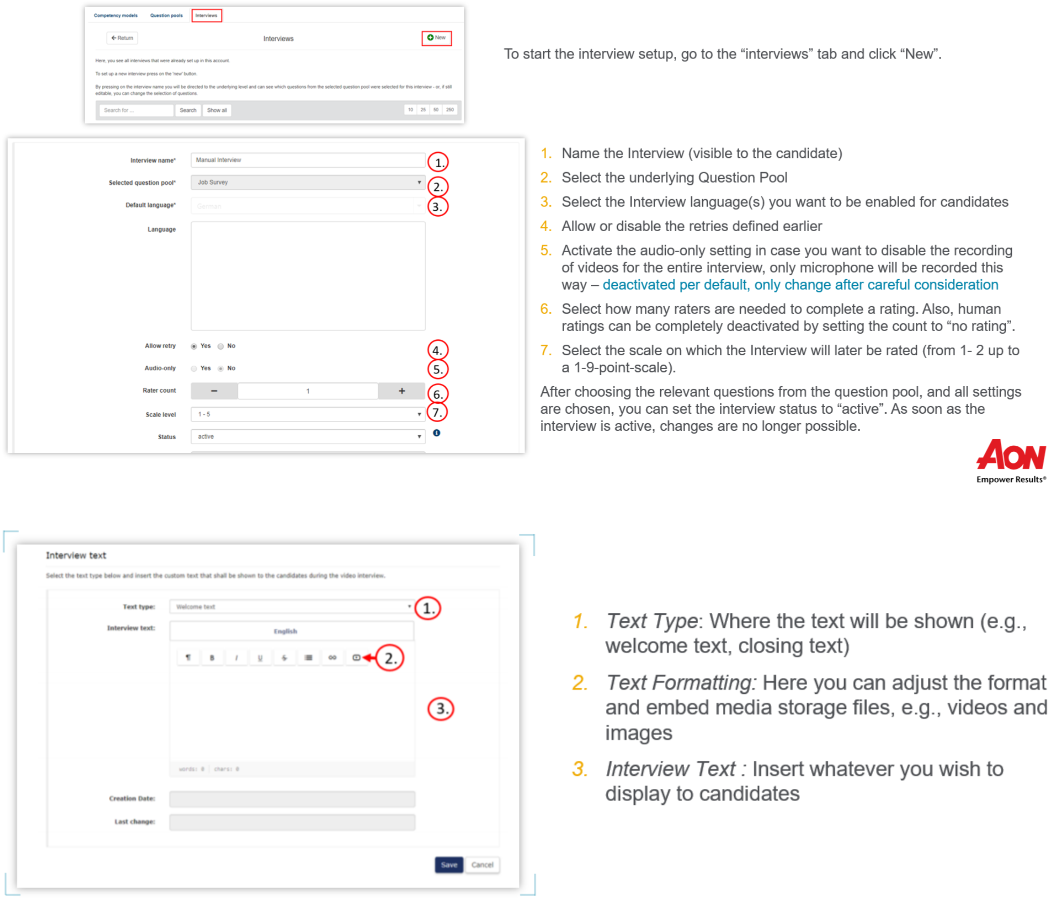
vidAssess is the Asynchronous video interview tool developed by Aon Assessment Solutions. It holds comparable features that can be found in similar tools such as those from HireVue, Sonru and ModernHire that are equally used in employment selection processes. vidAssess has been used in this work as the video interview collection tool and is, therefore, displayed in more detail. Given that it upholds the typical features of an Asynchronous video interview tool, those features are described in this chapter – not as a specialty of vidAssess itself but rather as a prototypical implementation of such standards. Therefore, any statements in this chapter that refer to vidAssess can be generalised to Asynchronous video interviews unless stated otherwise. This explicitly excludes the description of the automated scoring of video generated through vidAssess in the following chapter, which is unique to the vidAssess tool.

The general flow of vidAssess consist of four steps. The first step is for a company, recruiter or similar-placed interviewer to set up the Asynchronous video interview. The second step is for an interviewee to go through the process of completing vidAssess by generating respective videos. The third step is the rating of each of the videos, possibly combined with a compound scoring. Lastly, the fourth step is when the results are reported back to those involved.

The first step is the set-up of the interview (see Figure 5 for visualisation). It is completed on a system built to embody different customisation options. This way, questions can be generated specifically for a single process or used from a standard question library. The same applies to competencies or constructs that underline the questions and are to be rated for each of the videos. Different stimulus material can be chosen, such as welcome videos or outro sequences. Likewise, video, audio or text material can be chosen for questions to be displayed to each interviewee. Once those questions and competencies to be measured have been set up, the general flow of the interview is to be defined, including (but not exclusively) the overall timing (if fixed), timing per question, whether preparation time is allowed and whether or not retries and multiple attempts per question is allowed.

Lastly, within the first step, the rating process needs to be defined. For example, the number of observers per interviewee and how the allocation is carried out, whether or not open text options and comments are possible and rating scales are prominent examples that need to be defined for the rating process.

Figure 7. Example of screenshots describing the set-up process of vidAssess (slides 14 and 15 of Appendix II).



After the set-up, the second step, the video-generating process can be executed (see Figure 6 for visualisation). This process can be completed on any hardware and either via native apps or within a browser. It can also be completed without any restrictions to the person's location, as long as internet connection is available. Before each interviewee is presented with the specific questions, they undergo a preparation phase. During this preparation phase, all hardware aspects, such as the microphone and camera, are tested for proper functionality. The interviewee is tasked to complete a test recording and check if all aspects, such as lighting or other media properties, are to their satisfaction.

The existence of the preparation phase is especially interesting for this work given that it allows candidates to reflect how they record and further adjust any visual elements if desired.

Once the preparation phase has been completed, the different questions are shown one after the other and the interviewee is tasked to record an answer to each question. Before the video is uploaded and stored outside the hardware that is used for the recording, the interviewee can review and optionally reject to the upload to Aon's servers. This is to avoid any security or privacy breaches in case anything is visible or audible during the recording that is not intended by the interviewee. Again, this is explicitly mentioned given the relevance to visual cues that are available during the video recording and, thereby, shown by design or at least willingly accepted by the interviewee.

The third step consists of the rating phase which is typically completed by the recruiting company (see Figure 7). In rare cases, this is done by specialists that are hired for the purpose to rate video recordings. The borders between observer, rater and hiring manager will become very blurry during this step, as this could describe different roles and people or be the same person (even the same person who completed the set-up step).

Lastly, the fourth step is when the ratings are transferred into results and decisions are made based on these results. Any scores collected would be benchmarked against pre-defined expectations or specific pass marks. Typically, and vastly depending on the exact use of vidAssess, this would result in dividing the interviewees into different groups. Some will progress in a forward-moving way (such as passing on to the next step within a recruitment process or receiving a job offer) while others will not progress (such as not continuing with the primary focus of the process but either fading out or deviating the process's route).

Typically, different types of reporting are provided to different stakeholders of the vidAssess process. Narratives are typically used for those without a psychometric background, while score-based reports are often provided to HR admins and trained recruiters to have full access to their own interpretation options based on raw numeric data.

Figure 8. Example of screenshots describing the video-generating process of vidAssess (taken from slide 5 of Appendix II).

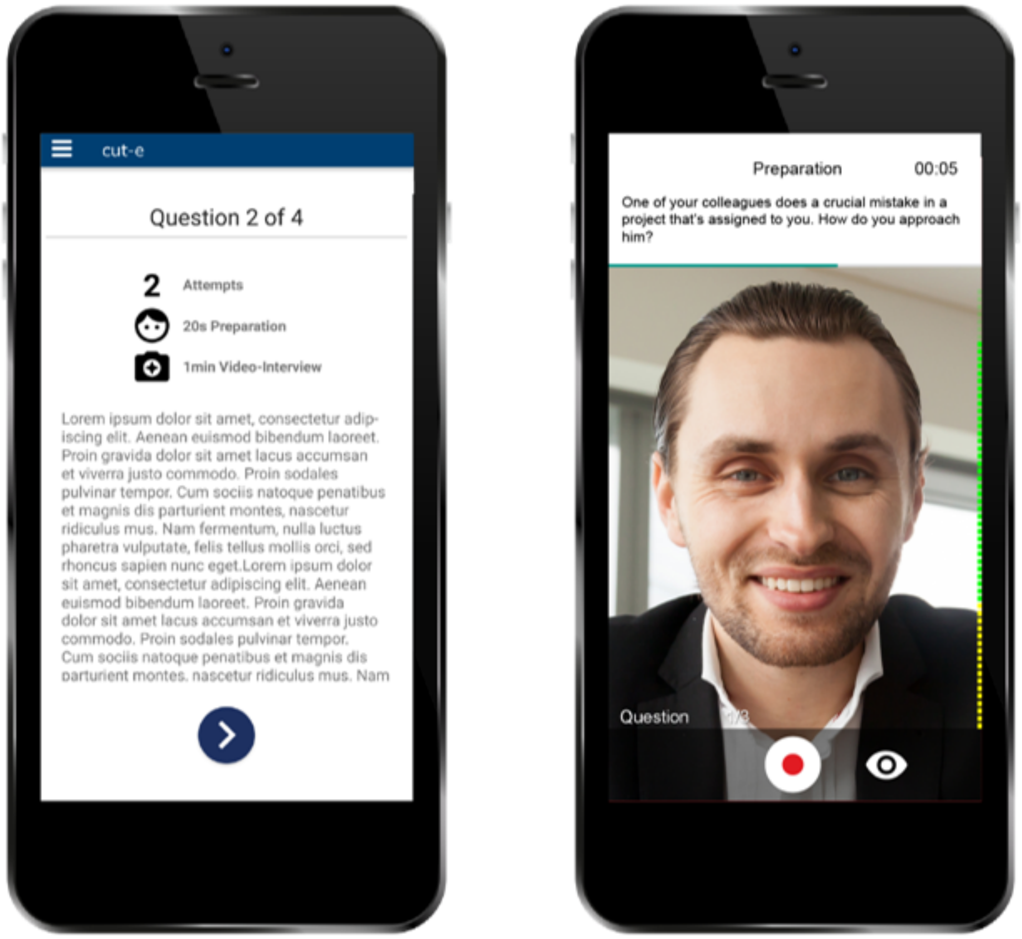
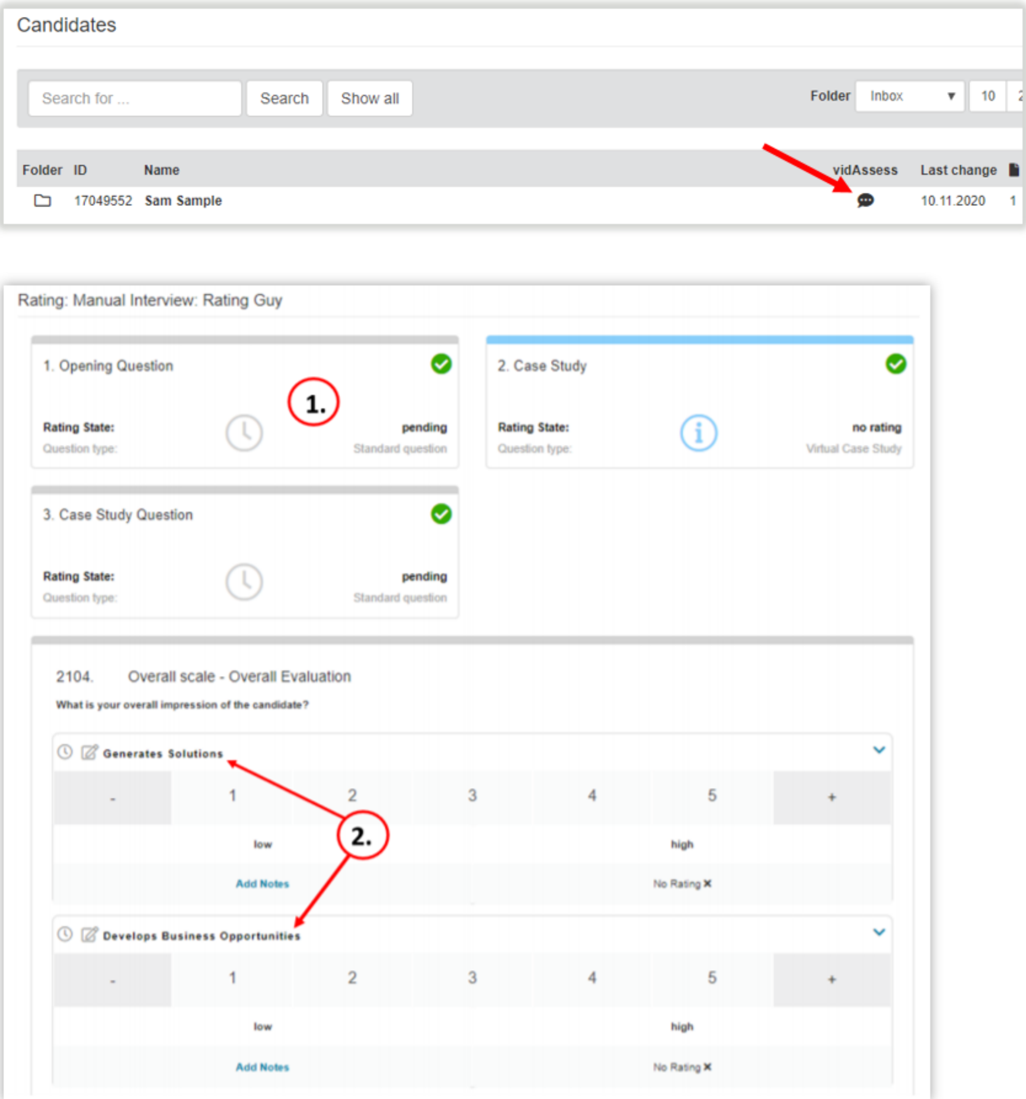


Figure 9. Example of screenshots describing the rating process of vidAssess (taken from slide 20 of Appendix II).



vidAssess AI Scoring

This chapter highlight the general functionality of how the automated scoring for vidAssess works. These scores are leveraged in the dataset for Study 3 and generating the scores is rather unique. This is why the scoring is explained superficially in this chapter. Further details can be found in the United States Patent Publication No US2021/0233030 titled 'Systems and Methods for Automatic Candidate Assessments in an Asynchronous Video Setting' by Preuss, Justenhoven, Kruse and Martin. If not referenced otherwise, this patent is leveraged for any statements that are made within this chapter. The patent can be requested through the US Patent Office or received from any of the mentioned inventors directly.

First the general flow and architecture of the data processing and scoring process will be described. Following this, relevant aspects of each of the steps will be explained in more detail. It should be mentioned upfront that automatic scoring does not have an impact on the video-recording process. The scoring process will only be done with videos that are already uploaded to Aon's server structure and, therefore, do not interfere with the recording directly.

For videos that are selected to be scored automatically, the first step is to split the audio and video data in order to isolate the audio information for further processing. Once split, an openly available speech-to-text service by IBM Watson is leveraged to generate a written transcript of what the interviewee said. This service is machine-learning based and trained with grammar rules, sentence structure and dialects to accurately use audio signals to detect words and form most likely word combinations – and ultimately full sentences. Each transcript is automatically enriched with an accuracy indicator, ranging from 0 (completely inaccurate) to 100 (completely accurate).

The next step is a Natural Language Classification model that is set up by Aon Assessment Solutions. It was developed by Preuss and Justenhoven; however, other inventors, not mentioned on the patent (Oke Brandt and Nico Tschöpe) contributed highly to the development of the model.

Natural Language Classification is one of the core tasks within the field of computational processing of natural language processing which, in turn, is core to the discipline of artificial intelligence (Gliozzo et al., 2017). Natural Language

Processing and, therefore, Natural Language Classification are to be viewed as being opposite to processing computational languages, such as C+ (Gliozzo et al., 2017). During this classification, the text is analysed and linked to predefined concepts in order to identify a text's topic, the literal meaning and the context it is placed in. The aim of a natural language classification is to derive the actual meaning of a text and also be able to fully automatically interpret (as a human reader would) how to enable fast decision-making on extensive volumes of written information (Indurkhya & Damerau, 2010). This process has been marked as a major future trend for selection processes (Lochner & Preuss, 2018).

At this point, it might be beneficial to dig deeper into general Machine Learning concepts; however; this would deviate too much from this work. Therefore, in the following work, the concepts of supervised versus unsupervised learning and, more specifically, dimensionality reduction and types of clustering within unsupervised learning are used without further explanation. However, *The Hundred-Page Machine Learning Book* by Burkov is excellent reading – especially when there are questions regarding some of the phrasing used in the following paragraphs.

The classification model that is used within vidAssess is a Bidirectional Encoder Representations from Transformers or BERT (Devlin et al., 2018). This leverages a combination of traditional convolutional neural networks and dimensionality reduction, where text is transformed into a vector representation.

One advantage of the used model is that – different to more classical networks – it takes into account the fact that the same word can have a different meaning depending on contextualised information, even within the same text or sentence (Devlin et al., 2018).

One example would be the word 'interesting'. While the answer to the question 'How was the university lecture?' can be 'interesting' with a very positive connotation, the exact same answer to the question 'How is the food?' might have a negative connotation.

The BERT algorithm is pre-trained to allow for multiple connotations and interpretations of the same word. This training is a two-step process, where the first step is happening unsupervised with access to a vast amount of text for the algorithm to understand and learn how words are embedded within a given form of text. The second step is a supervised learning task where the algorithm is taught to classify text to pre-defined classes. vidAssess scoring uses the underlying ADEPT-15 model, which has already been described.

Through human annotation, the algorithm is taught how to allocate information from the text to each class that represents an ADEPT dimension. In total, 30 classes are used, two per dimension that represent the far extreme for each dimension. This way, very positive indications (or rather positive) found in the text can be allocated to one class, while very negative (or rather negative) indications are allocated to the other class. Theoretically, the total amount of information per dimension is the combination of the two extreme classes per dimension.

Words, phrases and whole sentences can be allocated to one or multiple classes (and, therefore, to one or multiple ADEPT dimensions) and be linked in an either positive or negative way. This provides multi-dimensional information to one or multiple ADEPT dimensions at the same time.

The last step is how the positive and negative indications are used to form a score for each interviewee. The other relevant factor that is needed is the stimulus information. According to the inventors, the given information itself is not sufficient to derive the final scoring but can only be interpreted in relation to what stimulus was given, for typical cases, what question was asked.

Here is an example to illustrate this fact. Assuming a question is asked and that question is linked to the ADEPT trait Task Style with its dimensions' Structure and Drive, representing the conscientiousness trait in the ADEPT model. However, in this example, an interviewee has recorded a video as a response to the question and provided lots of positive indication to other ADEPT traits – but not to Task Style. This would be the equivalent of someone sitting in a traditional on-site interview and talking without actually answering the question.

This relationship between collected indication and provided stimulus triggers the thought that the scoring for each video is done by combining the sum of all positive and negative indications per dimension and relating it to the relevance of the trait that is linked to that exact question. The output of these scores are referred to as vidAssess AI scores and can be reported back in different aggregated ways. The most typical of these being: (1) 15 overall dimension scores combined across all videos provided by a single person; (2) per question scores for those dimensions that are pre-defined to be linked to the question and exclude any other data; and (3) reporting the raw negative and positive indication per question, allowing case-sensitive scoring adjustments in further processing steps.

Chapter 4: Studies

General Approach and Study Design Architecture

In this chapter, this project's studies are described. The details of each study evolved and were adjusted after findings of the previous study. However, the overall approach has been defined upfront and has been submitted as a working plan to the Freie Universität Berlin. The structure of the chapter follows the chronological flow of the studies. In the working plan, three separate studies are called out. After finishing the first study, reported in Jansen (2019) as well as in Jansen et al. (2020), it became clear that an additional follow up study for the first study was needed, potentially for the planned additional two studies as well. Therefore, moving forward, what has been described in the working plan as three studies will now be defined as three steps (Step 1, Step 2 and Step 3) of the methodological part of this work.

After finishing Step 1, more prominent changes have been discussed and put in place that changed the approach for the following two steps compared to what has been proposed in the working plan. Therefore, the overall aim of each of the three steps are summarised and described below for additional clarity and transparency.

In Step 1, systematic literature research was conducted to identify relevant visual cues for the asynchronous video interview setting. As a second source, two Thinking-out-Loud studies have been conducted specifically in the asynchronous video interview setting to generate additional visual cues. The list of cues did undergo various adjustments throughout Step 1, most of these to further condense and shorten the list of visual cues, as well as categorising and organising it thematically.

In Step 2, the list of visual cues is further processed to change it into an inventory. An inventory format is needed for the data-gathering phase in Step 3. Part of that change is a design layout that allows for clear and easy ratings, while watching asynchronous video interview responses. The other part is to identify the right cue/observer fit, meaning the optimal amount of cues an observer can rate without missing any. The outcome of Step 2 is a visual cue inventory rating sheet and subsequent instructions, as well as a specific procedure of how video interview responses are to be watched during the rating.

In Step 3, the main dataset is generated to contain all relevant variables such as, but not exclusively, self-rating, observer rating and systematic cue coding with the visual cue inventory. Different collection matrices and similar analyses are conducted on this dataset to gain clarity about the relationships between the different rating procedures and, therefore, being able to provide an answer to this work's research questions.

Table 5 displays an overview of the three steps with some general information for each step.

Table 5. Overview of the three empirical steps of this research work.

Step	**Study**	**Time Slot**	**Aim**	**Sample**
1	n/a	Jan to Mar 2019	generate visual cues based on literature research	n/a
	1	Mar to Aug 2019	generate visual cues based on ToL method	dataset 1
	1.5	Mar to Jul 2020	generate visual cues by tailoring ToL method to a more open question approach and include Thin Slices theory	dataset 1
2	n/a	Aug to Oct 2020	create a visual cue inventory	n/a
	2	Nov 2020 to Jan 2021	identify optimal coding procedure	dataset 1
3	3	Jan to Sept 2021	generate visual cue lenses and correlation matrices for final insights to answer initial research questions	dataset 2

Step 1: Generating the Visual Cue List

Step 1 consists of two parts: a literature review; and generating a visual cue list from this source – as well as the Thinking-out-Loud studies to enrich this list. Following the chronological order, extraction of visual cues from the literature is described first, followed by the two Thinking-out-Loud studies.

Reviewing literature, such as Gosling and colleagues (2002), Breil and colleagues (2021) and Mejia and Torres (2018), highlighted the need to define categories in which visual cues can be defined and, therefore, allows them to be searched for. Categories in that matter help to further funnel the literature review and identify research that produced cues which are specifically applicable for the asynchronous video interview setting.

A more-detailed overview as to which cue categories have been selected and the rationale for these is outlined in Chapter 2. The categories are Face, Body, Appearance, Media Properties and Environment.

Literature Review

Databanks and Sources

The databanks that have been used are PSYNDEX, PsycINFO, PsycBOOKS and PsycARTICLES (all have been accessed through EBSCO).

To generate a baseline which can be used for further processing, the literature review and, thereby, the identification of existing visual cues for each of the visual cue categories was done separately. During this process, visual cue lists were only considered from those publications that show a closeness to the research field of this work.

To summarise, the same 16 publications have been leveraged to form the visual cue list for the categories 'Face' and 'Body', given that most researchers have investigated those categories more holistically. However, given the previously described relevance and visibility of the face itself, this work will continue to treat the categories separately. Three publications have been leveraged to form the visual cue list for the category ‘Media Properties’. Four publications have been leveraged to form the visual cue list for the category ‘Appearance’. Given Gosling and colleagues’ (2002) and Gosling and colleagues’ (2005) previous research work, for the purpose of identifying environmental cues, the Personal Living Space Cue Inventory was included, as well as three additional publications. The full list of publications and their allocation to each visual cue category are in Appendix IV.

Defining Acceptance Criteria

As the next step in generating the visual cue list, three acceptance criteria were defined that each visual cue would need to meet to be included in this work’s visual cue list. Practically speaking, this often resulted in cues being rephrased in order to meet the acceptance criteria or, even more so, cues being removed (this is because cues that are already on the list are broad enough in their existing phrasing to fully overlap with the potential new cue[s]).

Due to the fact that the visual cues did undergo such intense changes throughout Step 1, additional clarity is needed and included by introducing a new terminology. From now on, all those cues that have not yet passed all acceptance criteria and are not yet named on the visual cue list will be called (visual) cue mentionings. Once they have passed all acceptance criteria – and likely changed in their phrasing or nature – they can be referred to as visual cues.

It must be disclosed upfront that, throughout the cue collection and list generation process, rephrasing and removing of cues happened at multiple points in time. Through the course of the upfront literature review, study 1 and study 1.5, as well as the point where the visual cue list was transformed into the visual cue inventory, adjustments to the cue lists have been made. Appendix XI is a graphical flow chart that summarises at which points in time these changes happened. The individual rationale for each of the changes will be further explained chronologically at the respective sub-chapter. Even though changes to the cue mentionings or the cue list have been made at multiple points in time throughout this work, the acceptance criteria that have been defined upfront have not changed and have always been the guiding principles for any changes made.

The aim of setting up acceptance criteria is to help generate a list of visual cues that is most relevant for the purpose of this work and, therefore, usable for later steps. Specifically, having in mind the data collection process where these visual cues are to be tracked and coded, similar to and in reference to Gosling and colleagues' (2002) data collection process.

The first acceptance criterion that each visual cue must meet is that it has to relate to a specific behaviour, movement or object or otherwise be phrased so that there is no need to further break it into multiple aspects to make it clearly observable. The terminology this acceptance criterion is further referred to in this work is 'Specificity'. Even though some visual cues that are put on the list may seem broad, they still relate to a specific aspect or area. Additionally, they would not need further explanation or allow for wide interpretation when coding the frequency or general presence in the videos.

The second acceptance criterion that each cue would need to meet is that is has to be controlled by the individual who is recorded in the video. Physiognomic or demographic cues are typical examples for areas of cues that would not meet this criterion and must, therefore, be removed from the visual cue list of this work. Making the assumption that the person being recorded had the free choice of the time and place of the recording, it can be further assumed that environmental cues or media properties (such as lighting) are influenced or chosen by the person in the video; hence under their control. One could argue that certain body or facial visual cues, such as sudden twitching or wild gestures cannot be controlled by those with respective disabilities. However, all videos used in the studies have been checked for visible disabilities and – if related to specific visual cues from the visual cue list – have been removed.

The third acceptance criterion that visual cues have to meet to make it onto the visual cue list for this work is that they have to be phrased so a coding can be carried out in a binary, linear way. Phrasing of cues, such as in the work of Nguyen (2015) and Batrinca and colleagues (2011), are problematic for the data collection phase of this work, as visual cues in their work contain multiple attributes that would not allow for bipolar coding that is targeted in this work. To meet this acceptance criterion, visual cues that were found through the literature review often had to be broken into multiple individual visual cues.

Data-driven Acceptance Criteria

Adding to the list of acceptance criteria, it also needs to be defined which data cleaning rules are applied, thus the data-drive acceptance criteria.[2]

The acceptance criteria are data-driven and, therefore, are not taken into account at this early conceptual stage within this work when applying them to the existing visual cue mentionings. However, for later reference and usage, they will be described in the following abstracts.

A significant number of papers dealing with the topic of cue reduction or cue elimination do not describe the data-driven acceptance criteria in detail (c.f. Back et al., 2011; Gosling et al., 2005; Mwangi et al., 2014; Qiu et al, 2015). If described, expert ratings or trained judges decide with cues to eliminate based on their relevance, redundancy or incidence.

To pick a specific example, Kasmar (1970) worked with bipolar pairs of adjectives to describe the general set up of rooms. Pairs have been excluded if the rated appropriateness of adjectives have been rated too low – even though the value has been set specifically, the rationale for that exact value has not been given. Likewise, if too many missing values have been present for a specific pair, it has been excluded – but the study did not specify why that specific value. In total, the excluding criteria for Kasmar have been:

- 75% of all ratings on appropriateness are above 7.5.
- Little variability among ratings (interquartile ranges of less than 3.0).
- No more than 3% missing values.

(Kamar, 1970).

[2]For discloser purposes, it needs to be highlighted that the data-driven acceptance criteria have only been derived after collecting dataset 2 and starting with the initial analyses work. Due to that factor, these criteria have not been applied during the Cue-Processing phase of Step 1 and Step 2.

Gosling and colleagues (2005) presented a multiple phase approach, where different raters reviewed their cues. However, no theoretical approach was given as to which cues were to be eliminated based on their descriptive statistics.

In conclusion, for this body of work, the thresholds as to which cues are to be excluded will be guided by the spirit of the articles mentioned above, as well as the expertise of the author and surrounding subject matter experts. The thresholds have been set rather inclusive and liberal when comparing with other numbers that could be found in mentioned articles. However, all defined criteria need to be met in order to be included for further analyses.

Firstly, each cue can only have 30% of missing cases, i.e., more than two-thirds of the ratings need to be distinct and explicit so that the cue is kept.

Secondly, too high intercorrelation between cues would be counter to the purpose of the analyses, i.e., if the overlap between two cues is too high, uniqueness and thus value, of the (second) cue diminishes. Therefore, cues that correlate with another cue of $r > |0.8|$ are to be excluded.

Thirdly, in order to display variance in personality traits, and doing so by using cues that link to different traits, zero or near-zero variance in a cue itself is problematic. Therefore, cues will be excluded if the ratio of the most common value to the second most common value is smaller than 98/2.

Any cues that do not meet either of the mentioned criteria will be excluded prior to any further analyses.

Applying Acceptance Criteria to Cue Mentionings from the Literature Review

Now that acceptance criteria have been defined, they can be applied to define which of the visual cues mentionings from the identified 23 publications are to be considered further for this work. A stepwise process is chosen, so that after each acceptance criteria step those cue mentionings that need rephrasing or adjustment can be remodelled before proceeding to the respective next acceptance criteria.

The 23 identified publications generated a combined list of 885 visual cue mentionings. After applying the first acceptance criterion, 223 visual cue mentionings are to be removed, keeping 662 cue mentionings for further processing. After applying the second acceptance criterion, 62 visual cue mentionings are to be removed, keeping 600 cue mentionings for further processing.

Even though the first two acceptance criteria resulted in the simple exclusion of visual cue mentionings, the third acceptance criterion resulted in the generation of additional cue mentionings. The generation of these additional cue mentionings are tracked in the same table and are labelled with a respective suffix. Existing visual cue mentionings from the literature review that contained too many options or multiple attributes were split so that each new visual cue mentioning only contains one specific attribute. Doing so, the overall number of visual cue mentionings increased by 102, having 702 cue mentionings for further processing.

Even though the following is strictly speaking not part of ‘applying acceptance criteria’, the list of visual cue mentionings contained an overlap. This is not unexpected, given that it is sourced from publications of similar manner and application. However, for the purpose of this work and to generate a visual cue inventory that can be used for systematic coding procedures, doublets or even partial overlap of visual cues are not desired. Therefore, any redundancies are to be marked and either cue mentionings being verbally merged together without violating the acceptance criteria or – specifically when dealing with full doublets – removing the redundant cue mentioning(s). In the following, this step is referred to as the fourth acceptance criterion. All four acceptance criteria are listed in Table 6.

Table 6. Overview of the four visual cue acceptance criteria.

	Acceptance criteria	Description
1	Specificity	This cue relates to a specific object, movement or behaviour.
2	Controllability	This cue can be controlled by the video respondee.
3	Binarity	This cue has to be represented in sets of binary oppositions.
4	Uniqueness	This cue has to be the only of its kind in the visual cue list.

The visual cue mentionings that are identified to be similar have been marked with an additional ID (temporary ID). For each group that is assigned with the same temporary ID, one or multiple new cue mentionings are created to both: (1) display the full spectrum of attributes of the visual cue mentionings from within this group but; (2) taking out any overlap of these newly-generated cue mentionings. The new cue mentioning(s) have been assigned temporary IDs as well in order to be able to track which original visual cues have been used as basis.

This step reduced the number of visual cue mentionings to 492, while also formally transforming them into visual cues, given that they meet all acceptance criteria and, therefore, form the starting point of the visual cue list.

However, the cues in the category 'Environment' so far included 365 items from the Personal Living Space Cue Inventory. Most are too detailed and specific for what can be observed during a video interview administration, such as types and genres of books that are on a bookshelf. Gosling and colleagues (2005) encourage a data-driven reduction of a cue inventory where needed. Therefore, of the 365 items in the Environment category, only 55 have been considered further. Table 7 shows the final numbers of visual cues per category.

Table 7. Number of visual cues after the literature research.

Visual cue category	Number of visual cues
Face	25
Body	47
Appearance	26
Media Properties	19
Environment	55

Dataset for Studies 1, 1.5 and 2

All three steps contain various studies and, in turn, each of the studies require a dataset containing asynchronous video interviews with which the participants of each study can work. To pursue this and future similar research work, the author initiated that two datasets were generated via the crowdsourcing marketplace, Amazon Mechanical Turk (mTurk). The first dataset is used for Studies 1, 1.5 and 2 of Steps 1 and 2 and is described in more detail in the below abstract. As the second dataset is solely used in Step 3, it will be mentioned and described there.

Dataset 1 was generated in March 2019 and contains a total number of 163 participants.

All participants had to complete a short questionnaire with biographical data, the ADEPT-15 questionnaire and had to respond to six questions via the vidAssess tool. The list of questions is displayed in Appendix V. As the video responses have been captured through the vidAssess tool, the previously-described vidAssess AI scoring produced ADEPT-15 scales for each participant. Total completion time was approximately 60 minutes.

From the original sample of n = 163, 21 have been excluded using general acceptance criteria related to the administration of the personality questionnaire, such as incomplete questionnaire administrations, low internal consistency (<.7) and too fast administration time (< four minutes) to exclude compromised datasets from any further steps.

Lastly, it is important to note that no data source from this dataset has been used multiple times. To be specific, all datasets that have been selected and incorporated for Study 1 are therefore excluded from Study 1.5. and Study 2. Furthermore, all data selected and incorporated for Study 1.5 are excluded from those that are selected for Study 2. This should minimise the risk of overfitting and wrongly confirming previously raised data-driven findings. Not all of the 142 video respondees have been allocated to a study, as the video material was greater than the number of available and needed participants for the upcoming studies.

Likewise, across the next few studies, a vast number of raters and coders had to be acquired to participate in the different experiments. Although all of them are employees of Aon, no employee actively rated or produced any sort of scores or content in more than one study. Again, this was done for the purpose of reducing any biases as much as possible in the data collection phase. Table 8 displays the biographical sample characteristics for dataset 1 in its reduced format (n = 142).

Table 8. Biographical details for dataset 1.

Area	Option	n	%
Highest Level of Education			
	Bachelor's degree	51	35.9%
	Some college, no degree	26	18.3%
	Associate's degree	19	13.4%
	Not Disclosed	15	10.6%
	High school graduate	10	7%
	Master's degree	8	5.6%
	Other (<5%)	13	9.2%
Currently Employed			
	Yes, full-time	80	56.3%
	Yes, part-time	24	16.9%
	No	22	15.5%
	Not Disclosed	16	11.3%
Industry of Employment			
	I am not currently employed	20	14.1%
	Not Disclosed	19	13.4%
	Commercial & Professional Services	15	10.6%
	Retailing	12	8.5%
	Consumer Services	11	7.8%
	Software & Services	10	7%
	Health Care Equipment & Services	8	5.6%
	Other (<5%)	47	33.1%
Ethnicity			
	White	97	68.3%
	Black or African American	14	9.9%
	Please select	13	9.2%
	Other (<5%)	18	12.7%
Gender			
	Male	64	45.1%
	Female	61	43%
	No active selection	15	10.6%
	Prefer not to Answer	2	1.4%

Study 1: Thinking-out-Loud

Having a complete and exhaustive list of visual cues to form the inventory from is crucial so that additional research steps can follow this work. Discussed upfront and backed by the findings of the literature review, such a visual cue list has not been collected for the specific setting (i.e., using asynchronous video interviews for selection purposes). Hence, in addition to the existing visual cue list that was derived from the literature, a second source of data is to be used that taps specifically into the setting of choice. That defines the data to be used, as asynchronous video respondees from candidates can be taken to generate additional visual cues. As a method of choice, the Thinking-out-Loud approach was chosen to generate additional visual cues and enrich the list that has been extracted from the literature review.

The Thinking-out-Loud approach allows for introspective insights (Bowles, 2010). It allows participants to focus on completing tasks while speaking out loud their internal processes and verbalising their perception, as well as potential interpretation layers that participants have formed while processing the tasks at hand (Ericsson, 2006).

The method is well suited for the approach of collecting additional visual cue mentionings in the specific setting as it does not intervene but rather supplement with watching videos. It also allows for direct access to what participants are focusing on and what they are perceiving. However, it needs to be distinguished between concurrent and retrospective Thinking-out-Loud procedures. (Van Den Haak, De Jong & Schellens, 2003)

Depending on the cognitive workload and the mental capacity that it takes to process the task at hand, at times participants are asked to complete Thinking-out-Loud procedures after completing the task itself (Charters, 2003). However, the task of producing additional visual cue mentionings from asynchronous video interviews is a comparably small cognitive load, especially as no other tasks are asked of the participants.

Pilot Study

A pilot study was set up to test the design of the planned study, specifically the instructions. For both the pilot study and Study 1, Maximilian Jansen was appointed study supervisor to then be able to incorporate respective findings in his maters thesis as well (see disclosure for more details). The pilot study was carried out with three participants (all Aon employees). The video material that was used for the pilot study was taken from dataset 1 as described before.

The study's design, specifically the instructions (Appendix VI) have been heavily influenced by Bowles (2010). The instructions consisted of a definition of visual cues, examples of visual cues, as well as operationalised, verbalised descriptions of visual cues that are desired, as well as negative examples. Both English and German mentionings are taken into account, where English is the preferred language if it does not increase the cognitive workload.

The specific task directed at the participants is to watch the videos and say aloud visual cues of the person providing the video, in line with the definition and description given at the beginning. To further reduce cognitive workload, the design of the study allowed participants to solely focus on one of the five categories of visual cues as defined in this work. With that, participants were watching a total of six videos: the first was intended as a warm-up session, to get comfortable with the setting and allow participants to get used to the task. During the remaining five videos, each participant was asked to focus on one category per video in the following order: Non-Verbal-Face; Non-Verbal-Body; Environment; and Media Properties.[3]

Each of the three participants from the pilot study watched the six videos from one randomly-selected video respondee of the mentioned sample, in order to maximise video material that was used while not confounding participants with completely different video material, showing completely different people for each of the cue categories. The videos were presented muted and with only visual information. All sessions were audio recorded and have been transcribed to infer the mentioned

[3]At the time of Study 1, the categorisation and naming of the visual cue categories as displayed in Chapter 2 had not been completed yet, hence the difference to the ones used in Study 1.

visual cue mentionings. After capturing visual cue mentionings, all participants were asked to provide feedback, specifically with respect to the instructions and general study design.

The pilot study's aim was to ensure the study's design, specifically that instructions are working well to ensure additional visual cue mentionings can be captured during the Thinking-out-Loud process in the desired way. Therefore, those cues captured during the pilot study are not taken into account. Moreover, the interesting outcome is the feedback that was provided after the visual cue-capturing phase.

All three participants of the pilot study confirmed that the instructions are clear and understandable. From their perspective, no information or details are missing to complete the task of capturing visual cues when watching the muted videos.

During the actual task, all participants started to call out cue mentionings; however, there are four findings that are interesting to highlight and that will have an impact to the procedure of the actual Thinking-out-Loud study, respectively. First, all participants noted the same visual cue mentionings more than once, especially when they detected no further visual cue mentionings. Secondly, all participants added explanations to the visual cue mentionings they found, which was not asked nor needed for the purpose of the visual cue generation. Thirdly, all participants moved away from the definition of cues that were given in the beginning and made generic as well as interpretative impressions instead of actual observable cues. Fourthly, clearly visible cue mentionings have deliberately not been mentioned by participants, because they were asked to focus on one specific category – cue mentionings that fall in any of the other four categories have been ignored.

To tackle the first three findings, a list of prompts is introduced that the study supervisor can use to direct participants' attention to a specific area. 'Was there anything else you took into account for your rating?' or 'Did you notice anything else in the video?' are two example prompts that can be given to participants to direct their attention to focus on generating visual cues aligned with the definition given (see Appendix VIII for full study instructions). The other change that was made to the instructions is that participants are now free to name any visual cue mentionings, ignoring any categorisation to ensure clearly present cue mentionings are not neglected. They are also asked to provide a relevant rating at the end of the video, solely based on the observed visual cue mentionings. This is to further stimulate participants to try and observe as many different visual cues as possible and to simulate to a normal rating of asynchronous video interviews and, therefore, potentially generate cues that increase the cue utilization within the Lens Model.

Study 1

Study 1 was carried out with 10 participants (all Aon employees).

The sample consisted of five men and five women (50% each). The average age was 25 years (SD = 1.7). As per the areas of education, most (n = 8, respectively 80%) had a psychology background (either general, business or educational), the remaining two participants studied business management. Five participants (50%) held a master's degree or equivalent university degree, three (30%) a bachelor's degree and the remaining two (20%) were without university degree. Eight (80%) of the participants were German, two (20%) were United States citizens. Four (40%) for the participants indicated that they had previous experience in conducting interviews.

As the sample of videos (dataset 1) is far greater than the available participants of the study, a subset of videos is to be selected and will be included in Study 1. One option is to only use a specific video (for instance, the first) from each person that submitted videos. However, looking through the videos and backed by reports from Brenner (2019), one can observe how people who submitted their videos are getting more comfortable with the setting over time. Therefore, the hypothesis was raised that different visual cues might be present, depending on a video respondee's level of comfortableness. In addition, each video respondee answered the same six questions; however, visual cues might change, depending on which question they are answering. In conclusion, both these findings led to the decision to include all six videos for each video respondee.

To use a variable that is not related or dependent of the actual study, the first 60 video respondees have been selected for Study 1, corresponding to a total of 360 videos (six videos per video respondee).

Taking into account previously-mentioned order effects and to maximise exposure of video respondees to participants, a specific sorting system was used. This ensured that, while participants are watching a maximum of six videos each: (1) every video is watched only once; (2) every participant watches each video respondee only once; and (3) every participant watches videos that have been recoded to answer different questions.

Given external constraints, the time limit for each participant in this study was set at approximately 30 minutes. Depending on how long it took for instructions and warm-up, as well as duration of the individual videos that participants watched, not every participant was able to work through all six videos. The minimum number of videos that have been watched is three. A total of 54 videos have been processed in this study.

Different to the pilot study, no warm-up task was included as the pilot study strongly indicated that all participants understood the instructions correctly; however, also because prompts have been added to the study supervisor's arsenal to guide participants' behaviour. In addition, and different to the pilot study, the study supervisor was asked by the author to take notes directly. This ensured that any questions, exact phrasings and confirmation that is needed to fully align with the intention of the participants can be done either right after the study exercise or even during the capturing of the visual cue mentionings. This was especially relevant as the preferred language still remains English, which is not the native tongue to all participants. Some even chose to verbalise partially or fully their observations in German. The study's instructions can be reviewed at Appendix VIII.

The highest total number of cues mentioned by a participant was 126, the lowest number of cues that was mentioned was 37 (SD = 24). Each participant watched five videos (SD = 1) on average and mentioned 11 cues (SD = 6) while doing so. This generated a total of 609 visual cue mentionings that have been gathered through the Thinking-out-Loud approach in Study 1.

Visual cue mentionings that have been produced in German (n = 88) are translated into English; however, these are marked separately with a respective suffix in case further research shows that this has an effect and these visual cue mentionings need to be reworked.

Applying Acceptance Criteria to Cue Mentionings from Study 1

The same four acceptance criteria that have been applied to the visual cues list generated from the literature review are now applied to the collected list of visual cue mentionings from Study 1.

Applying the first acceptance criterion, 53 visual cue mentionings are removed, keeping 556 for further processing. After applying the second acceptance criterion, 14 visual cue mentionings are to be removed, keeping 542 cue mentionings for further processing.

Four of the transcripted cue mentionings had to be reworked to meet the third acceptance criterion. Even though the aim – to capture and transcribe the visual cue mentionings during the study in a way that they already meet this third acceptance criterion – is slightly missed, the difference to the far greater number of visual cue mentionings that had to be reworked from the literature review shows how well the transcription was executed.

As to meet the fourth acceptance criterion, the list of visual cue mentionings, specifically doublets, had to be reworked so that, in total, 416 cue mentionings were removed from the final list, resulting in a total of 130 visual cues generated through Study 1.

Combining Visual Cue List from Literature and Study 1

Before undergoing the first critical review of the process so far, the two lists are merged, resulting in a single list of 317 visual cues. However, given the two distinct sources of data, once again the fourth acceptance criteria need to be applied, reducing the number of cues to 243 by removing 59 doublets. The final numbers and the split across the different categories are shown in Table 9.

Table 9. Number of visual cues of literature review and study 1.

Visual Cue Category	Number of Visual Cues
Face	38
Body	58
Appearance	43
Media Properties	35
Environment	69

Emerging Thoughts Driven by Study 1 and their Impact to the Next Steps

In the order of events (according to the proposed working plan), the combination of the literature review and the Thinking-out-Loud approach should have been the foundation of the visual cue inventory list. This would be further processed in the upcoming steps. However, after Study 1 was completed, several discussions emerged between the author and those who have significantly contributed to the success of this work, specifically with Krumm, Preuss, and Jansen.

The key topics that have been discussed are:

(1) Both the pilot study and Study 1 showed the highest number of cue mentionings within the first third to half of a vidAssess response. Anything mentioned in the remaining time are more likely to be repetitions and did not help to increase the overall list of visual cues. To increase the number of cue mentionings collected, participants were encouraged to only note cue mentionings and not explain them. While this aligned with the overall approach and purpose of the first study, information on why certain cues were mentioned or considered relevant was missed.

(2) The set-up and procedure of Study 1 appeared appropriate and easy to understand for both participants and researchers.

(3) The decision to omit a warming-up task in Study 1 appeared to not have significantly impacted task performance or data quality and was not mentioned by participants.
(4) The rating task appeared to be helpful. No participant of Study 1 mentioned that the task felt artificial or strange, compared to all participants of the pilot study (without rating task) expressing such feelings.
(5) The videos selected as stimuli for Study 1 were balanced regarding their position within each candidate's interview, based on assumptions regarding possible changes in behaviour throughout the interview process. How this affected the distribution of personality aspects and, therefore, the cues generated in Study 1 were not accounted for.

Although clear agreement has been reached that Study 1 has been majorly successful in providing additional and most likely relevant visual cues, the discussed topics also showed that there are some improvements and learnings that can be transferred if the study was to be re-done. The author took the decision to re-run the Thinking-out-Loud study and account for at least some of these implications – most of all, as the visual cue list is the essence of this work. If significant numbers of visual cues are left uncovered and are not being included in the upcoming steps, any conclusions or follow ups, be it within this work or from research fellows, will show limitations. Therefore, the additional effort has been taken to run another study, from here on to be called ‘Study 1.5’, with a slightly-adjusted design to gather further visual cues in the respective specific setting. Study 1.5 has been reported in Lüders (2020) and Lüders acted as the test supervisor for the data generation phase, including the pilot study.

Study 1.5: Thinking-out-Loud (Refactored)

Given the five topics that have been discussed in the aftermath of Study 1, the following implications have been taken into account for the design of Study 1.5:

(1) Use of thin slices – only the first minute of video recordings – instead of the whole video. Research on thin slices supports their use and validity for interpersonal judgements (Ambady & Rosenthal, 1993; Brenner, 2019).

Study 1 indicates diminishing returns beyond thin slice lengths of videos. Reducing the length per video also allows for more videos to be watched by each participant and, respectively, even more opportunities to generate additional visual cue mentionings.

The use of thin slices in the context of interpersonal judgements based on vidAssess responses will be a valuable addition to research on asynchronous video interviews and may result in further research avenues on candidate behaviour and its changes throughout the interview process, as well as effects on interviewer judgements.

(2) As part of a shift towards personality and cue personality associations, Study 1.5 will use a more open response format, encouraging descriptions of how and why impressions have been formed. This will require recording participants which would allow for qualitative data analyses if required in a separate work.

(3) The overall procedure will be similar to that of Study 1. Omitting a warm-up task is expected to have an even lower impact on data quality, due to the higher number of videos per participant.

(4) A rating is to be incorporated again to make the procedure more accessible and guide the Thinking-out-Loud task.

(5) Video responses to be used as stimuli will be balanced regarding their position within the interview process again. However, in an additional step, the data will be reviewed regarding the distribution of personality traits based on ADEPT-15 ratings. If necessary, data will be manually selected to ensure normal distribution of the ADEPT-15 aspects.

As can be seen in (2) and (5), a key aspect of Study 1.5 is a shift of focus towards personality ratings. To account for the impact of personality attributes to the presence of visual cues, video responses for Study 1.5 are preselected so that the full range of different personality traits are covered in the videos. To accomplish this, remaining data from dataset 1 is selected.

For each of the 15 dimensions that the ADEPT-15 questionnaire reports out, the normal distribution is visually checked to ensure the sample for Study 1.5 includes video respondees with the full spectrum of personality attributes, in case any visual cues are heavily related to a specific attribution (for example, only appear if a person is very high on extraversion). Based on a visual check, all dimensions appeared to be normally distributed enough so that the sample seems to be good enough to progress further (see Appendix IX for the distribution per dimension).

Pilot Study

As some potentially significant aspects have been changed for the new study design, once again a pilot study is being carried out to ensure specifically that the instructions are well received and the design is fitted to produce a high number of cue mentionings.

The pilot study is conducted with two participants, both SME employees at Aon. The video material was taken from the same cleaned mTurk sample that was used in the main study; however, not from the preselected 60 video respondees.

Both participants are given upfront instructions to introduce the task, which is to watch muted videos, and to then rate two defined personality traits and verbalise which visual information they have used to make that judgement. Lastly, they have been asked to explain why the respective visual information was relevant for the rating and what their general impression of the video respondee is.

The pilot study surfaced a few learnings for the main study that are reported in the following. The respective changes that have been drawn and which impact the design and, specifically the actions, of the test supervisor of Study 1.5 are picked up in the following sub-chapter.

(1) Both participants spent much time (>5 minutes) getting familiar with the operationalised anchors given to each personality trait. The purpose of the operationalisation was to help participants to understand the correct definition of each trait, respectively what a high or low value on each trait would intend. In addition, both participants stayed relatively close to their description of visual cue mentionings and, at times, even fully quoted words from the trait description as visual cue mentionings.

(2) Even though the pilot study consisted of only two participants, they preferred to make the personality trait rating and visual cue mentionings in different order. One participant reported that it was easier to first collect the visual cues and then come up with a rating while the other participant reported the opposite.
(3) Linked to (2), it became obvious that juggling two personality traits and their respective ratings and links to visual cues can become confounded if not clearly stated by either the setting or proactively by the participant which visual cue mentioning refers to which trait.

Design and Procedure of Study 1.5

Firstly, the pilot study's learning needs to be incorporated to adjust the design of the study. The instructions and the provided material have been adjusted to address: (1), most of the operationalised anchors that describe the personality traits have been removed, only some are left to ensure the traits' meanings are still understood. A few specific examples of good visual cue mentionings are added for more clarity. This further helped to remove dependencies during the study to the more generic anchors and shifted the focus to specific visual aspects of each video.

To address: (2), the study advisor leaves it to each participant's choice whether they report the visual cue mentionings first and complete the ratings based on their summary, or the other way around.

However, in line with learning: (3), each trait is discussed individually and separated, so that there will be a least possible mix up of visual cue mentionings and respective trait(s). In addition, more specific prompts are added to the test supervisor's arsenal, so they can guide the participants during the rating process and optimise the number of visual cue mentionings.

To ensure clarity with this work's reader, the design and execution of Study 1.5 is summarised in the following.

Study 1.5 is carried out with 10 participants (70% female) and, like all previous studies, the participants have been Aon employees. All participants have been equipped with material where the personality traits from the ADEPT-15 model to be rated are outlined, together with an operationalised description. They are also equipped with rating sheets for these personality traits to fill in their ratings. Furthermore, they are instructed to watch six videos, whereas each video is used only once during the whole study. The videos show only the first 60 seconds and have been muted. Once watched, the participants are then instructed to rate the personality traits on a five-point Likert scale to which the video is a response and describe the visual cue mentionings they have used to reach their rating decision. Rating and speaking out visual cue mentionings can be done in the order of the participant's choice. As the last step per trait rating, the participants can further elaborate on why the cue mentionings are connected and relevant to the respective trait.

The use of a more open response format and recall prompt is intended to yield data, allowing for more insight into why certain cues have been deemed relevant. Learnings from Study 1.5 will inform and hopefully trigger subsequent research on cue-trait associations.

The trait rating and calling out cue mentionings is done twice per video for the two distinct traits that the video respondees are responding to. Given the ADEPT-15 model consists of six traits, each trait is covered twice in the study design. Together with the fact that each video is used only once in the whole study, these two aspects allow for maximum spread of stimuli and, therefore, ideally result in the highest possible number of cue mentionings and ultimately visual cues.

Applying Acceptance Criteria to Cue Mentionings from Study 1.5

Being able to leverage the existing visual cue list, the cue mentionings are compared to what the list already entails. Any mentionings that are already included in the list can be neglected from further processing; even if they are transformed into proper visual cues, they are redundant to the existing ones on the list. This comparison excluded 489 of the 570 cue mentionings, reducing the cue mentioning list down to 81 for Study 1.5.

Applying the first acceptance criterion, four visual cue mentionings are removed, keeping 77 for further processing.

All 77 cues passed the second acceptance criterion and no cues had to be removed at this stage.

Ten of the transcripted cue mentionings had to be reworked to meet the third acceptance criterion, adding to a total of 87 cues after the third acceptance criterion.

In order to meet the fourth acceptance criterion, the list of visual cue mentionings, specifically doublets, had to be reworked. Therefore, in total, 50 cue mentionings were removed from the final list, resulting in a total of 37 visual cues being generated through Study 1.5. This also includes removing doublets that are already included in the existing visual cue list formed from the literature review and Study 1. Table 10 shows the overview of number of cues per category that has been derived through the various methods of Step 1.

Table 10. Visual cues of step 1.

Visual Cue Category	**Number of Visual Cues**
Face	52
Body	60
Appearance	49
Media Properties	38
Environment	81

Step 2: Designing the Visual Cue Inventory

To further proceed and assess whether or not visual cues are present in a video, the cues cannot be kept in the list. Raters would need a more digestible set-up, otherwise the risk of missing out cues and not being able to track them properly are too high. An inventory needs to be set up that meets several design and psychometric characteristics, so that tracking visual cues is simple, easy and reduces errors to a great extent. This tracking will be called ‘systematic cue coding’ or, in short, ‘coding’.

Differentiation between Dynamic and Static Cues

The visual cues on the list have all been checked for the third acceptance criteria; namely to ensure the coding procedure can happen in a linear way. At the starting point of developing the cue inventory, the cues are split into those called ‘dynamic cues’ and ‘static cues’. This does not change the allocation to the categories, but rather adds a label that helps to decide how best to set up the coding procedure. Static cues do not change their attributes during the course of the video. The expectation is that, even the smallest subset of a video (respectively a photo or screenshot of the video), would reveal all static cues. On the other hand, dynamic cues could occur at different points during the video in all variety of frequencies.

This additional sorting has a lasting impact to the overall number of visual cues (as shown in the following example).

In the category ‘Appearance’, there are currently multiple cues attributed to one’s hair. While these can all be coded in a binary way, some are exclusive to one another and exchangeable to a certain degree. Some cues can be paired up to form the extreme ends of a scale, such as ‘dark hair’ versus ‘light hair’ or ‘tended hair’ versus ‘untended hair’. Working through the visual cue list in this way, the overall number of cues that need to be individually coded in the systematic cue-coding process is reduced to 236.

Overall, there are two distinct ways of coding dynamic versus static cues, which is why this differentiation is so relevant. In total, of the newly-sorted 236 cues, 95 are labelled as dynamic cues and the remaining 141 are static cues.

This is the last step of cue processing within this work. Technically, this should have been completed within Step 1, but the adjustment for some of the cues (as well as the split into dynamic and static cues) are to be counted as relevant steps in the processing flow. All further steps from this point will not affect the actual visual cues, but rather the structure of the inventory and the procedure of capturing said cues.

Closing the cue-processing approach of this work, Figure 10 summarises all steps and puts them into order and perspective. To make it easier to read, the same figure is additionally attached to Appendix XI in a larger scale. In addition, Table 11 shows the final overview of number of cues per category for this body of research, including their percentages per category.

Figure 10. Cue-processing flow.

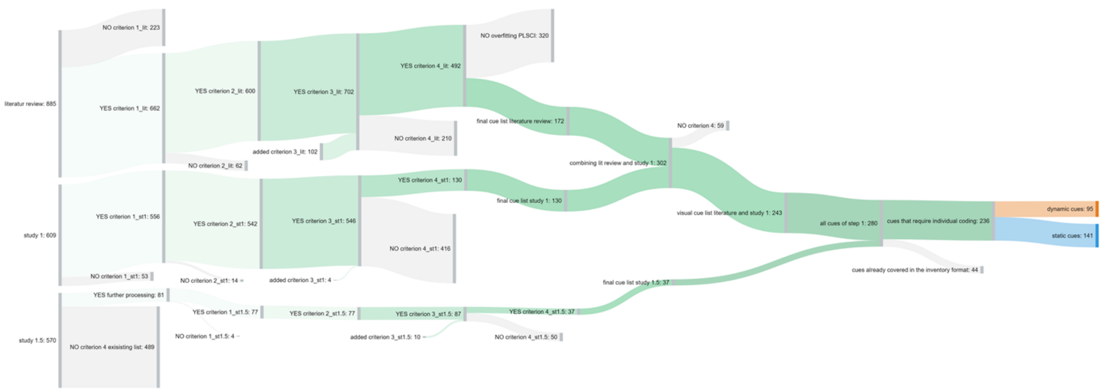


Table 11. Final overview of visual cues.

Visual Cue Category	Number of Visual Cues	Percentages
Face	39	17%
Body	55	23%
Appearance	43	18%
Media Properties	30	13%
Environment	69	29%
Total number of cues	**236**	

Structure and Order of the Visual Cue Inventory

Next, the inventory's look and feel are tackled to optimise the coding procedure. Leveraging best practice from design-thinking approaches and guidelines from Eppler and Kernbach (2016), small icons are added to each area of cues. This allows the eyes to be guided when looking for a specific cue that was visible during a video. Likewise, cues that show relative closeness are grouped and displayed together, such as cues for the forehead, eyes and mouth. In the same manner, cues for the head are displayed 'on top' of arms and shoulder cues and cues related to hands are displayed before those that relate to fingers and nails. Static cues are separated out and displayed in a separate part of the inventory, as these can always be coded in a one-off manner.

For all dynamic cues, a line is added to each cue, on which the actual frequency can be tracked. This tally sheet helps to externalise the counting process for the coders. Afterwards, they can, based on the tally sheet, fill out a standardised Likert scale to indicate the relative frequency. This two-step tracking is helpful when looking at dynamic cues where different absolute frequencies are expected.

A good example might be the cues 'smiling' and 'duck face'. Specifically given the setting in which the videos are generated, one would expect the cue smiling to be present more often than the cue duck face.

A general counting transformation that stays the same for all dynamic cues would overemphasise cues that have a higher expectancy rate. Therefore, coders will be asked to first fill out the tally sheet and to then transfer it to the Likert scale, based on the experience they have of rating video respondees in general and coding specifically.

The only outstanding discussion is on the levels of the Likert scale, as well as the operationalisation. After thoughtful discussion, it was decided to use a four-point Likert scale. This will allow for enough freedom to differentiate between each level, while also keeping the scale comparable to the interindividual specific interpretations. The operationalisation is chosen to be ‘none, minimal, some and a lot’ for all dynamic cues for consistency and, again, easier handling of the coding.

For the static cues, the same four-point Likert scale is introduced, reducing errors in further data aggregation and analyses. However, given the static cues are too different, instead the two extreme sides per cue are operationalised and tied to ‘1’, respectively ‘4’. The middle levels of the scale do not have a specific description but act as anchors for coders to indicate a tendency to one of the sides per cue. In addition, a fifth coding option per static cue is introduced when the cue cannot be rated (n.a./‘5’). However, some static cues do not perfectly allow for the bipolar scale option, which is why they are displayed as grouped multiple choice options. A good example would be the type of worn make-up which lists the options of ‘made-up eyes’, ‘rouged lips’ and ‘highlights’. As any combination of the previous cues can be present, they need to be individually coded. The same is true for a range of other static cues, which are, therefore, presented in the same manner.

The general set-up follows basic design principles, such as those reported in Eppler and Kernbach (2016). Font size, type and layout of each page are in line with those, but also very pragmatic aspects, such as handleability during the coding procedure, influenced the layout. The final template of the visual cue inventory is attached to Appendix X.

Study 2: Defining the Maximum Number of Cues that can be Coded

The next step of ordering the visual cue inventory is to determine the maximum number of dynamic cues that can be coded while watching a video. This is specifically relevant for the data generation phase. On the one hand, the coding needs to be done in a very efficient way in order to not misuse participants' time. On the other hand, having to code too many cues at the same time will result in missing out some of them. Having to remember hundreds of cues, looking for them constantly in the video and coding and filling out the tally sheet will be too much of a load for the participants. Therefore, the ideal threshold needs to be determined where cues can be coded in a sufficient way, while keeping the number of cues that are coded simultaneously as high as possible. As indicator of choice, Cohens' kappa (Cohen, 1960) is used to establish the level of agreement between all coders.

Hendrix, who reported parts of these findings in her work (2021) as well is appointed to be the study supervisor for Study 2. Nine additional employees of Aon are recruited as participants.

Of the 10 participants of Study 2, seven are female (70%) and three are male (30%). The average age was 25.5 years (SD = 2). As per the educational level, eight (80%) hold a bachelor's degree and two (20%) hold a master's degree. Six (60%) of the participants have a psychology background, while the others are unique in their studies. Likewise, six (60%) are German, two (20%) are Indian and the other two are unique to their country of origin.

For the purpose of measuring the intercoder level of agreement, an adaptive trial-and-error approach is leveraged. This means that the starting point of how many cues are included in the first round is defined. The next round(s) will be dependent on whether the level of agreement is sufficient (i.e., the number of cues can be increased) or the level of agreement is not sufficient (i.e., the number of cues need to be reduced). At the start, 50 cues are set to be coded in parallel, as this is the number that is tracked in in-person group exercises or role plays as well (Obermann, 2018). In addition, coding just five cues is included to be able to generate an intercoder agreement benchmark.

Study Design and Procedure for Study 2

As described before, an adaptive trial and error is used to generate the ideal number of cues coded in the upcoming step (Step 3). This will require multiple rounds, though each round will be run in the same manner.

(1) For each round, three lengths of videos are used that represent the typical lengths of the videos being generated in this setting: 60 seconds; 120 seconds; and 180 seconds.
(2) For each round, the study supervisor will generate a ground truth baseline, watching all videos as often and with as much fast forward and rewinding as needed to ensure all visual cues per round are tracked once to the fullest extent.
(3) Each video will be watched by two randomly-chosen participants, so that intercoder agreements can be measured.

Given the above three statements and the number of cues to be observed for the first round to be five and 50, a total of six videos are needed (see Table 12).

Table 12. Video-coder allocation for study 2, round 1.

Video	Length in Seconds	# of Cues Rated	Coding
video 1	60	5	ground truth
video 1	60	5	rater 3
video 1	60	5	rater 4
video 2	120	5	ground truth
video 2	120	5	rater 1
video 2	120	5	rater 3
video 3	180	5	ground truth
video 3	180	5	rater 2
video 3	180	5	rater 4
video 4	60	50	ground truth
video 4	60	50	rater 1
video 4	60	50	rater 2
video 5	120	50	ground truth
video 5	120	50	rater 2
video 5	120	50	rater 4
video 6	180	50	ground truth
video 6	180	50	rater 1
video 6	180	50	rater 2

Comparing the intercoder agreement for video 1 to video 3, where five cues have been coded by two participants per video, the level of agreement is nearly perfect (mean $\kappa = .86$) based on Landis and Koch (1977). On the other hand, the intercoder agreement for video 4 to video 6 still reveal substantial agreement (mean $\kappa = .67$) when coding 50 cues. As these findings drive the decision for the second round for Study 2, it becomes clear that choosing 50 cues as a starting point, informed by Obermann (2018), has been likely a good anchor for the systematic cue coding as well.

Therefore, the next round will be around those 50 cues to find out how the intercoder agreement changes when increasing and respectively decreasing the number of cues. In the second round, the same procedure will be done as in the first round, this time with 40 and 60 cues per coding. Again, videos of three different lengths will be used to ensure the length of the video – which is a variable with respective range and typically high standard deviation – does not have an impact in the coding procedure, such as fatigue (for longer videos) or single missed opportunities (for shorter videos) while coding. The coding procedure was conducted by the remaining five participants.

Expectantly, the increase to 60 cues per coding procedure decreased the intercoder agreement to a moderate agreement (mean $\kappa = .45$). However, decreasing the number of cues to 40 per coding procedure did not meaningfully increase the intercoder agreement (mean $\kappa = .69$).

Moreover, for video 4 to video 6, the mean values of the coders per video correlated in a highly significant way to the ground truth (mean $r = .93$). These two findings combined show evidence that, not only do the raters agree among themselves, but nearly all cues have been coded correctly through the systematic coding process when the number of cues are close to 50.

Looking at the visual cue list with a pragmatic eye, one can quickly see that these results mean to split the dynamic cues into two groups, that are to be coded individually. Given similarities of content and type of cues, section A of the cue inventory will consist of 43 cues to cover image quality, face and leg area. Section B of the cue inventory will consist of 52 cues to cover all other body-related cues.

At the last step, the static cues need to be tackled and checked to see if any issues arise to have them coded as a single block. For that, a third round of Study 3 is introduced. Here, once again, six randomly-picked videos are leveraged from the early described sample and are only controlled for the overall timing to be close to 60 seconds, 120 seconds and 180 seconds for two videos each. The difference to the previous rounds was that, this time, all coders are allowed to pause, rewind and fast forward, given all cues are present throughout the video anyhow by definition of being static. As expected, all coders showed the natural tendency to watch the videos for a few seconds, pause the video and fill out the coding sheets. The average intercoder agreement across all six videos shows a substantial agreement (mean $\kappa = .69$).

This result highlights that: (1) the natural chosen tactic of coding static cues has proven to be very successful and more so that; (2) all 141 static cues can be coded in one session. Therefore, only one other section needs to be added to the final visual cue inventory and this will be called section C and will contain all 141 static cues.

For any further use of the visual cue inventory, both within this work and in additional research, it is recommended to have every coder complete all three sections per video (but in three different sessions). This allows for a fully-completed visual cue inventory administration per coder, while maximising the quality of data these generate during the coding process.

As a reminder, the final visual cue inventory, including the earlier described graphical and visual adjustments as well as the split into three sections, is displayed in Appendix X.

Step 3: Generating the Dataset and Answering Research Questions

This step is the final one of this work to answer research questions. So far, the previous steps enabled the opportunity to do so. Firstly, a list of visual cues was generated through systematic literature research and Thinking-out-Loud studies. Next, a visual cue inventory was formed from this and a specific coding procedure was defined for it. Now, within this step, the visual cue inventory is used to help produce a dataset that allows upfront defined questions to be answered.

Method

A sample, including video material and self-rating, is generated through the mTurk platform. Trained observers are asked to provide personality judgements based solely on the video material. Additionally, coders are systematically tracking all observable visual cues for all video material. Lastly, the previously-mentioned natural language classification algorithm is used to provide automatically-generated personality ratings for each of the video respondees.

Dataset 2 for Study 3

The dataset that is leveraged for this study has not been used in previous studies of this work. It was generated via the mTurk platform and follows the same procedure as the data collection for dataset 1. The dataset was produced in October 2019.

Each participant was first presented with a demographic survey, followed by a job advertisement for a trainee position (see Appendix XII). They were asked to imagine that they were applying for the job as real applicants. After reading the job advertisement, participants provided answers to each of the six interview questions (see Appendix XIV). Instructions for each interview question emphasised that a work-relevant example was required (see Appendix XIII). After the video recording, participants took ADEPT-15 and then the Mini-IPIP. Total assessment time was approximately 60 minutes.

In total, n = 140 datasets were generated through the mTurk platform; however, following the same guidelines as described for dataset 1, 41 datasets had to be screened out so that the final dataset included n = 99 video respondees. Table 13 displays the biographical sample characteristics for dataset 2.

Table 13. Biographical details for dataset 2.

Area	**Option**	**n**	**%**
Highest Level of Education			
	Bachelor’s degree	40	40.4%
	Some college, no degree	16	16.2%
	Master's degree	12	12.1%
	Associate's degree	11	11.1%
	High school graduate	6	6.1%
	Other (>5%)	14	14.1%
Currently Employed			
	Yes, full-time	60	60.6%
	Yes, part-time	25	25.3%
	No	9	9.1%
	Not Disclosed	5	5.1%
Industry of Employment			
	Retailing	16	16.2%
	Consumer Services	14	14.1%
	I am not currently employed	8	8.1%
	Software & Services	7	7.1%
	Health Care Equipment & Services	6	6.1%
	Technology Hardware & Equipment	6	6.1%
	Not Disclosed	5	5.1%
	Commercial & Professional Services	5	5.1%
	Other (>5%)	32	32.3%
Ethnicity			
	White	60	60.6%
	Black or African American	12	12.1%
	Asian	10	10.1%
	Asian	9	9.1%
	Not Disclosed	5	5.1%
	Other (>5%)	3	3%
Gender			
	Male	48	48.5%
	Female	45	45.5%
	Prefer not to Answer	6	6.1%

Additions to Dataset 2

To complement the reduced dataset of n = 99, three subject matter expert employees from Aon have been asked to produce observer ratings for both the ADEPT-15 traits and the developed third-person version of the Mini-IPIP (as described in the respective chapter about the IPIP). One of the observers provided the scores for the ADEPT-15 traits, the other two observers completed the Mini-IPIP questionnaire for each mTurk worker.

In the previous chapter, the link to how this work can contribute to additional automation in processing asynchronous video interview has already been established. For this purpose, additional information is collected for the dataset that will likely be relevant in subsequent work (but is not used in this work). Shortly summarised, for each video the dataset contains auxiliary technical information, such as word count, character count and automatic confidence scores provided by the used Speech-to-Text Microsoft Azure API. This API also produced a transcript for each video, which is then further processed by Aon's vidAssess AI algorithm to produce ADEPT-15 trait scores for each video through a natural language classification process. The detailed description about how the vidAssess AI algorithm is working can be found in the respective chapter about the vidAssess AI scoring. These scores will be included in some of the following analyses, specifically when it comes to comparing the results of self-ratings and observer ratings.

As previously said, all the video auxiliary data will not be further used to answer the research questions but will enrich the dataset for additional work in this space and hold a promising follow up to any results that may be generated by answering research questions.

Lastly, a systematic coding (with the help of the visual cue inventory) is added to the dataset. For this, a total of 11 coders (73% female, age mean = 24.5, SD = 1.7) have been extensively trained to use the visual cue inventory in the previous outlined way. This process includes that all six videos from any mTurk worker are reviewed by the same coder and each video is reviewed a total of three times (once each for sections A, B and C) by that coder. All 11 coders are employed by Aon. Table 14 displays their educational level and familiarity with interview ratings.

Table 14. Biographical details for coders of dataset 2.

Area	Option	n	%
Highest Level of Education			
	Bachelor’s degree	9	81.8%
	Master's degree	2	18.2%
Current Role in Aon			
	Intern	4	36.4%
	Working student	4	36.4%
	Associate	2	18.2%
	Consultant	1	9.1%
Experience with Interview Data			
	Yes	4	36.4%
	No	7	63.6%
Experience with Conducting Interviews			
	Yes	3	27.3%
	No	8	72.7%

For transparency, it needs to be disclosed that the number of mTurk workers that each coder reviewed and coded did vary. This is mainly due to the fact that this work was completed on a voluntary basis and it was to the discretion of each individual coder how much time they were willing to spend on this project.

Results

Descriptive statistics

The dataset contains multiple datapoints that can be checked to ensure a normal and expected data structure. The total numbers already take into account cleaning out if required per the previous mentioned requirements.

Table 15 shows details for the IPIP and the ADEPT-15 questionnaires, both the self-rating and the observer rating, as well as the ADEPT NLC ratings. For readability purposes, the table and all following tables that contain FFM trait descriptions only shows the initial letter of the respective personality trait (i.e., 'O' for 'openness', 'C' for 'conscientiousness', 'E' for 'extraversion', 'A' for 'agreeableness' and 'N' for 'neuroticism').

For readability purposes, the descriptive statistics for the visual cues that were completed by the subject matter experts are split into five parts, following the cue categorisation logic. Tables 16 to 20 show the respective details. Table 21 shows the list of 59 cues that have been excluded from the previous table and thus, any further analyses based on the previously-described cleaning rules. As a reminder in short, this was due to high missingness of data (> 30% of cases missing), very high intercorrelation with another variable ($r > |0.8|$) or zero or near-zero variance (ratio of the most common value to the second most common value < 98/2).

Table 21 shows relevant technical feature statistics highlighting the general video data quality available in dataset 2. Appendix XV contains all those features on a question-by-question level for an even deeper insight into data structure. Word Count, Character Count and STT Confidence are values that have been produced by the IBM microservice that has been used to generate the ADEPT NLC scores. The audio and video quality ratings have been manually rated by the observers on a 0 to 5 scale, where anything below '2' would have been too poor to be used.

Table 15. Descriptive Statistics for Self- and Observer Ratings.

		N	M	SD	Min	Max	Skew	Kurtosis
IPIP Self-ratings	O	99	9.89	2.57	4.00	16.00	0.53	-0.60
	C	99	11.86	1.92	8.00	17.00	0.54	0.22
	E	99	11.08	2.08	5.00	18.00	0.22	0.69
	A	99	11.93	1.89	7.00	18.00	0.22	1.15
	N	99	10.74	2.20	7.00	18.00	0.48	0.55
ADEPT Self-ratings	O	99	0.15	0.31	-0.87	0.85	-0.30	0.31
	C	99	0.19	0.32	-0.57	1.37	0.47	1.01
	E	99	0.04	0.38	-1.25	0.80	-0.64	0.31
	A	99	0.11	0.26	-0.63	0.72	-0.11	-0.18
	N	99	0.14	0.28	-0.68	0.81	-0.42	0.37
IPIP Observer ratings	O	99	13.05	2.33	7.00	18.50	-0.15	-0.31
	C	99	15.78	2.02	9.50	20.00	-0.48	0.29
	E	99	14.18	2.18	8.00	18.50	-0.22	-0.56
	A	99	13.87	1.91	9.50	18.00	-0.21	-0.67
	N	99	12.40	0.83	9.50	14.00	-0.12	0.50
ADEPT Observer ratings	O	99	3.02	1.05	1.00	5.00	0.27	-0.73
	C	99	3.34	1.03	1.00	5.00	-0.16	-0.71
	E	99	3.27	1.08	1.00	5.00	-0.19	-0.70
	A	99	3.27	0.97	1.00	5.00	0.11	-0.80
	N	99	3.21	0.95	1.00	5.00	-0.22	-0.92
ADEPT NLC ratings	O	99	0.25	0.12	-0.10	0.48	-0.78	0.98
	C	99	0.41	0.19	-0.14	0.92	-0.24	0.38
	E	99	0.30	0.13	-0.09	0.59	-0.73	0.54
	A	99	-0.04	0.13	-0.33	0.25	0.03	-0.71
	N	99	-0.18	0.12	-0.56	0.12	-0.34	0.93

Table 16. Descriptive Statistics for visual cues, category: **Face**.

visual cue	n	M	SD	Min	Max	Skew	Kurtosis
Friendly expressions	99	2.49	0.78	1	4	-0.14	-0.75
Cheerful expressions	99	1.84	0.76	1	4	0.53	-0.47
Interested expressions	99	2.70	0.78	1	4	-0.32	-0.63
Self-assured expressions	99	2.62	0.85	1	4	-0.11	-0.89
Timid expressions	99	1.42	0.58	1	4	1.59	2.91
Calm expressions	99	2.77	0.73	1	4	-0.06	-0.50
Surprised expressions	99	1.41	0.53	1	3	1.16	0.24
Sceptical expressions	99	1.64	0.58	1	3.5	0.84	0.38
Arrogant-amused expressions	99	1.20	0.45	1	3	2.29	4.35
Grumpy expressions	99	1.31	0.56	1	3.83	2.09	4.31
Indifferent expressions	99	1.61	0.80	1	4	1.25	0.58
Strong expressions	99	1.89	0.80	1	4	0.41	-1.02
Diverse facial expressions	99	1.66	0.73	1	4	0.91	-0.02
Rapidly changing expressions	99	1.53	0.69	1	4	1.20	0.69
Wrinkled forehead	96	1.91	0.83	1	4	0.57	-0.76
Looking at camera	99	3.12	0.76	1	4	-0.48	-0.60
Looking sideways	99	2.24	0.81	1	4	0.22	-0.77
Looking up	99	1.98	0.86	1	4	0.41	-1.04
Looking down	99	2.20	0.96	1	4	0.24	-1.24
Rolling eyes	97	1.56	0.83	1	4	1.34	0.55
Heavy blinking	97	1.55	0.71	1	3.33	1.02	-0.35
Wide open eyes	97	1.93	0.88	1	4	0.56	-0.86
Closed eyes	97	1.45	0.61	1	3	1.10	0.02
Narrow eyes	97	1.80	0.84	1	3.83	0.81	-0.56
Tinkling with eyelashes	97	1.17	0.45	1	2.83	2.65	5.74
Furrowed eyebrows	98	1.72	0.84	1	4	1.00	-0.10
Raised eyebrows	98	2.32	0.88	1	4	0.12	-1.10
Laughing	99	1.30	0.56	1	3.17	1.92	2.53
Smiling	99	1.75	0.70	1	3.67	0.73	-0.43
Biting lips	99	1.21	0.52	1	3.17	2.52	5.08
Licking lips	99	1.42	0.59	1	3.17	1.32	0.59
Pressed lips	99	1.68	0.76	1	4	1.17	0.84
Plopping lips	99	1.34	0.67	1	4	2.10	3.83
Duckface	99	1.15	0.44	1	4	4.03	18.92
Pout	99	1.09	0.25	1	2.33	3.03	8.71
Wide open mouth	99	1.60	0.81	1	4	1.28	0.61
Pausing	99	2.14	0.70	1	3.83	0.41	-0.48
Swallowing hard	99	1.72	0.71	1	3.67	0.94	-0.10
Fast mouth movements	99	2.01	0.92	1	4	0.33	-1.20

Table 17. Descriptive Statistics for visual cues, category: **Body**.

visual cue	n	M	SD	Min	Max	Skew	Kurtosis
Straight posture	83	2.61	0.92	1	4	-0.10	-1.05
Slouched posture	83	2.22	1.02	1	4	0.33	-1.18
Relaxed posture	89	2.75	0.88	1	4	-0.27	-0.95
Maintained posture	93	2.98	0.83	1	4	-0.35	-1.02
Open posture	85	2.94	0.83	1	4	-0.31	-0.99
Stiff posture	89	1.65	0.66	1	3.33	0.81	-0.48
Body tilted	95	1.60	0.78	1	4	1.44	1.62
Body turned away	94	1.23	0.54	1	3.67	2.54	5.94
Parallel body orientation	86	3.24	0.84	1	4	-0.97	0.17
Leaning backward	96	1.48	0.64	1	4	1.42	1.56
Leaning forward	95	1.78	0.78	1	4	0.80	-0.21
Sitting	81	3.86	0.47	1	4	-4.13	18.48
Standing	81	1.02	0.08	1	1.5	3.98	16.46
Walking around	99	1.02	0.11	1	1.8	6.05	37.80
Swivelling on chair	96	1.32	0.69	1	4	2.50	5.57
Change of position	97	1.30	0.54	1	3.8	2.37	6.09
Head tilt	99	1.89	0.80	1	4	0.55	-0.74
Head pulled back	99	1.38	0.60	1	3.33	1.44	0.87
Head oriented away from camera	99	1.55	0.72	1	3.67	1.22	0.36
Sloped head posture	98	1.48	0.60	1	3.33	1.05	0.26
Straight head posture	98	2.98	0.79	1	4	-0.65	-0.01
Nodding	99	1.72	0.83	1	4	1.14	0.40
Shaking head	99	1.71	0.77	1	4	1.34	1.33
Shoulder movement	82	1.93	0.80	1	4	0.73	-0.25
Touching head	99	1.30	0.53	1	3.5	2.32	5.14
Touching face	99	1.60	0.76	1	4.5	1.66	2.56
Touching hair	99	1.26	0.54	1	4	2.75	8.13
Touching body	87	1.35	0.55	1	3.5	1.83	3.05
Adjust clothing	85	1.13	0.33	1	2.6/	2.80	7.43
Holding hands in front of mouth	72	1.22	0.50	1	3.33	2.39	5.06
Biting nails	85	1.05	0.18	1	2	3.97	15.39

Table 18. Descriptive Statistics for visual cues, category: **Appearances**.

visual cue	n	M	SD	Min	Max	Skew	Kurtosis
Good personal cleanliness	97	3.30	0.83	1	4	-0.92	-0.07
Colourful clothes	99	1.85	1.01	1	4	0.88	-0.47
Dark clothes	99	2.69	1.09	1	4	-0.20	-1.30
Distinctive clothes	99	1.57	0.81	1	4	1.16	0.20
Shiny clothes	98	1.63	0.84	1	4	1.07	0.10
Wrinkled clothes	83	1.83	0.82	1	4	0.44	-1.06
Much skin visible	94	1.62	0.78	1	4	1.17	0.90
Business-like clothes	91	1.74	0.99	1	4	1.09	-0.07
Athletic clothes	89	1.97	1.12	1	4	0.73	-0.94
Tank Top	94	0.05	0.23	0	1	3.92	13.50
T-Shirt	94	0.43	0.50	0	1	0.30	-1.93
Button-up shirt/blouse	94	0.22	0.42	0	1	1.31	-0.30
Pullover	94	0.19	0.40	0	1	1.54	0.39
Loose-fitting top	70	2.46	0.86	1	4	0.06	-0.70
Hair cover face	97	0.23	0.42	0	1	1.29	-0.35
Asymmetrical haircut	82	0.32	0.47	0	1	0.77	-1.42
Plucked eyebrows	88	0.32	0.47	0	1	0.77	-1.43
Tended hair	86	2.85	1.06	1	4	-0.35	-1.19
Dyed hair	82	0.15	0.36	0	1	1.97	1.88
Dark hair	89	3.10	1.09	1	4	-0.88	-0.63
Distinctive hair	86	1.78	1.03	1	4	0.95	-0.52
Long hair	79	2.10	1.25	1	4	0.48	-1.49
Heavy make up	91	1.26	0.57	1	4	2.37	5.91
Made-up eyes	94	0.12	0.32	0	1	2.35	3.54
Rouged lips	94	0.04	0.20	0	1	4.46	18.09
Highlights	94	0.02	0.15	0	1	6.53	41.07
Glasses	99	0.41	0.50	0	1	0.34	-1.90
Jewellery	94	0.11	0.31	0	1	2.51	4.36
Earrings	92	0.08	0.27	0	1	3.15	7.98
Necklace	98	0.04	0.20	0	1	4.57	19.09
Nose piercing	98	0.05	0.22	0	1	4.02	14.30
Tattoo	91	0.02	0.15	0	1	6.41	39.57
Headphones	92	0.09	0.28	0	1	2.88	6.39
Headset	92	0.10	0.30	0	1	2.66	5.15
Clothing item	99	0.10	0.30	0	1	2.61	4.85

Table 19. Descriptive Statistics for visual cues, category: **Media Properties**.

visual cue	n	M	SD	Min	Max	Skew	Kurtosis
Person in centre	99	0.84	0.37	0	1	-1.81	1.29
Light from left side	95	0.36	0.48	0	1	0.58	-1.68
Light from right side	95	0.26	0.44	0	1	1.06	-0.89
High resolution image	99	2.75	0.89	1	4	-0.02	-0.96
Complete person and background visible	99	0.06	0.24	0	1	3.63	11.27
Only Head/face visible	99	0.30	0.46	0	1	0.84	-1.30
Complete face visible	99	0.83	0.38	0	1	-1.72	0.95
Only parts of face	99	0.02	0.14	0	1	6.72	43.57
Eyes visible	98	0.88	0.33	0	1	-2.27	3.18
Hands visible	98	0.18	0.39	0	1	1.61	0.60
Gestures visible	98	0.36	0.48	0	1	0.59	-1.67
Big distance to camera	99	1.79	0.59	1	3	0.09	-0.47
Big distance to wall	96	2.46	1.10	1	4	0.06	-1.35
Camera level with head	99	0.59	0.50	0	1	-0.34	-1.90
Camera above head	99	0.10	0.30	0	1	2.61	4.85
Vertical picture frame	99	0.06	0.24	0	1	3.63	11.27
Dark light	99	2.16	0.85	1	4	0.37	-0.49
Artificial light	90	2.66	1.26	1	4	-0.18	-1.64
Even light	98	2.83	1.06	1	4	-0.38	-1.12
Light halo	99	0.14	0.35	0	1	2.03	2.13
Face lit by screen	98	2.40	1.01	1	4	-0.02	-1.17
Overexposed	96	0.69	0.47	0	1	-0.80	-1.38
Light from behind	95	0.21	0.41	0	1	1.40	-0.05
Camera shaking	99	1.66	0.90	1	4	1.11	-0.02

Table 20. Descriptive Statistics for visual cues, category: **Environment**.

visual cue	n	M	SD	Min	Max	Skew	Kurtosis
Another person present	97	0.02	0.14	0	1	6.64	42.57
Living room	77	0.26	0.44	0	1	1.08	-0.86
Kitchen	77	0.08	0.27	0	1	3.09	7.64
Bedroom	77	0.26	0.44	0	1	1.08	-0.86
Public location	89	0.07	0.25	0	1	3.39	9.62
Plain room	84	3.02	0.99	1	4	-0.56	-0.92
Tidy room	76	3.08	1.03	1	4	-0.66	-0.91
Finished wall	86	0.41	0.49	0	1	0.37	-1.88
Painted wall	82	0.29	0.46	0	1	0.90	-1.21
Wall pattern	87	0.13	0.33	0	1	2.21	2.92
Window	92	0.17	0.38	0	1	1.69	0.88
Blinds	92	0.12	0.33	0	1	2.31	3.36
Shutters	91	0.10	0.30	0	1	2.64	5.04
Curtains	92	0.09	0.28	0	1	2.88	6.39
Door	80	0.30	0.46	0	1	0.86	-1.28
cookware & pots	96	0.02	0.14	0	1	6.61	42.07
bed	98	0.10	0.30	0	1	2.59	4.75
Sofa	98	0.12	0.33	0	1	2.27	3.18
chair	98	0.13	0.34	0	1	2.13	2.58
desk	97	0.04	0.20	0	1	4.54	18.84
table	97	0.04	0.20	0	1	4.54	18.84
drawer	97	0.09	0.29	0	1	2.76	5.70
file cabinet	97	0.09	0.29	0	1	2.76	5.70
Wardrobe/closet	98	0.09	0.29	0	1	2.78	5.81
Shelve	98	0.16	0.37	0	1	1.79	1.23
Book shelve	97	0.08	0.28	0	1	2.99	7.01
Stereo stand	97	0.02	0.14	0	1	6.64	42.57
crate	97	0.02	0.14	0	1	6.64	42.57
coat rack	97	0.02	0.14	0	1	6.64	42.57
Finished wall	97	0.11	0.32	0	1	2.40	3.80
Painted wall	97	0.16	0.36	0	1	1.88	1.55
Mirror	97	0.03	0.17	0	1	5.34	26.74
Calendar	97	0.02	0.14	0	1	6.64	42.57
Book	98	0.14	0.35	0	1	2.01	2.06
Collection	98	0.03	0.17	0	1	5.37	27.08
Map	98	0.02	0.14	0	1	6.68	43.07
Toy	98	0.04	0.20	0	1	4.57	19.09
Stationary	97	0.03	0.17	0	1	5.34	26.74
Toilette	98	0.03	0.17	0	1	5.37	27.08
Bag	98	0.03	0.17	0	1	5.37	27.08
Lamp	98	0.09	0.29	0	1	2.78	5.81
Fan	98	0.03	0.17	0	1	5.37	27.08
Electronic equipment	98	0.12	0.33	0	1	2.27	3.18
Athletic equipment	98	0.02	0.14	0	1	6.68	43.07
Weapon	98	0.02	0.14	0	1	6.68	43.07
Decoration	98	0.18	0.39	0	1	1.61	0.60
Plant	98	0.02	0.14	0	1	6.68	43.07
Pillow	99	0.07	0.26	0	1	3.30	8.97

Table 21. Descriptive Statistics for visual cues, **cleaned out.**

category	visual cue	n	M	SD
Appearance	Groomed appearance	95	3.14	0.91
Appearance	Rolled-up sleeves	22	0.14	0.35
Appearance	Trousers	2	0.50	0.71
Appearance	Loose-fitting bottom	2	3.50	0.71
Appearance	Tied-up hair	48	2.21	1.40
Appearance	Groomed beard	33	2.70	1.10
Appearance	Specialized clothing item	99	0.00	0.00
Appearance	Jewellery in the background	99	0.00	0.00
Environment	Any pet present	97	0.01	0.10
Environment	Big room	58	2.47	1.08
Environment	Open-spacious room	63	2.41	1.06
Environment	Carpet	27	0.11	0.32
Environment	Fridge	96	0.04	0.20
Environment	Ice cube machine	96	0.00	0.00
Environment	food	96	0.00	0.00
Environment	nightstand	98	0.02	0.14
Environment	Garbage can	97	0.01	0.10
Environment	tie rack	97	0.00	0.00
Environment	Wall pattern	97	0.00	0.00
Environment	Clock	96	0.01	0.10
Environment	Magazine	98	0.04	0.20
Environment	CD/record	98	0.01	0.10
Environment	Game	98	0.00	0.00
Environment	Musical instrument	98	0.00	0.00
Environment	Medication	98	0.00	0.00
Environment	Tool	98	0.01	0.10
Environment	Label	98	0.00	0.00
Environment	Religious artifact	98	0.01	0.10
Environment	Appropriate decoration	28	0.68	0.48
Media Properties	Camera below head	99	0.33	0.47
Media Properties	Complete person visible	99	0.01	0.10
Media Properties	Only upper body visible	99	0.71	0.46
Media Properties	No parts of person visible	99	0.16	0.37
Media Properties	Person in lower half of frame	99	0.11	0.32
Media Properties	Horizontal picture frame	99	0.94	0.24
Body	Shaking legs	6	1.58	0.92
Body	Legs crossed	6	1.33	0.82
Body	Legs open	6	1.33	0.82
Body	Crossing arms	49	1.26	0.54
Body	Arms behind back	48	1.12	0.36
Body	Crossing arms behind back	48	1.13	0.33
Body	Open arms while sitting	52	2.44	1.06
Body	Rubbing hands together	47	1.37	0.65
Body	Empathetic gesture clapping	48	1.25	0.56
Body	Circling hands	56	1.68	0.87
Body	Waving	52	1.30	0.58
Body	Folding hands	43	1.46	0.74
Body	Symmetrical hand position	39	1.47	0.69
Body	Hands resting on table	36	1.54	0.79
Body	Hands cling to something	40	1.41	0.84
Body	Fidgeting/gesturing with object	39	1.21	0.56
Body	Interacting with device	40	1.35	0.68
Body	Fast movement	64	2.00	0.86
Body	Slow movements	63	1.69	0.63
Body	Big gesture	68	2.04	0.91
Body	Insulting gesture	52	1.03	0.14
Body	Pointing	68	1.24	0.42
Body	Rubbing fingers	49	1.20	0.45
Body	Twisting fingers	48	1.24	0.54

Table 22. Descriptive statistics for technical features of videos.

	N	M	SD	Min	Max	Skew	Kurtosis
Word Count	97	311.82	76.62	89.33	425.5	-0.58	-0.38
Character Count	97	1272.08	315.29	400.67	1653.67	-0.58	-0.55
STT Confidence	97	76.69	9.22	55.33	89.83	-0.50	-0.87
Audio Quality	92	4.09	1.35	0	5	-1.67	1.90
Video Quality	92	4.36	1.26	0	5	-2.40	4.99

In total, the displayed descriptive statistics show normal behaviour so that one can proceed with looking at the further correlative results. As a reminder, in addition to the visual cues that have been cleaned out (Table 21), there have been 41 datasets cleaned out from dataset 2 already. That is, 41 participants were dropped in the initial cleaning stage due to having not met the predefined data quality criteria and subsequently, a total of 59 cues were removed based on the described cleaning rules.

Order of Results

The results are displayed in a systematic way that allows the reader to explore the different aspects of the Lens Model. Each component of the Lens Model is looked at individually, first the cue validity, then cue utilization and finally the functional achievement. This will happen on an aggregated level in the form of regression values and vector correlations. Thereafter, specific and detailed findings will be added on a more granular level, such as correlations on individual cue level. Lastly, the results of the different studies – including those from Step 1 and Step 2 – will be linked and sorted to the research questions.

Cue Validity

In Brunswik's Lens Model, cue validity refers to the link between the true underlying construct being evaluated (e.g., extraversion) and the cues that form the lens. In the absence of an absolute truth for the target traits, self-ratings of personality by the video interviewees are used as the criterion. As outlined in Chapter 3, two personality measures are used – ADEPT-15 and IPIP. Both provide measures of the Big 5 personality traits, against which cue validity is evaluated.

Ten multiple linear regression models were run to test if the visual cues significantly predicted self-reported personality traits. Only visual cues that were significantly correlated with the target trait were included in the models. Results of regression models using Big 5 as measured by IPIP are reported first, followed by models using ADEPT-15. All details to the below regression models can be found in Appendix XVI. They are listed in the same order as they are mentioned in the below text.

IPIP

Model 1 – Openness. Multiple linear regression was used to test if visual cues significantly predicted self-ratings of openness. Eleven visual cues were included in the model. The overall regression was statistically significant (R^2 = 0.49, F(11,72) = 6.36, p < .001). It was found that hair covering the face (β = 0.29, p < .01), earrings (β = 0.31, p < .01), the presence of a map (β = 0.21, p = .02), and heavy blinking (β = -0.25, p = .03) significantly predicted self-perceptions of openness.

Model 2 – Conscientiousness. Multiple linear regression was used to test if visual cues significantly predicted self-ratings of conscientiousness. Five visual cues were included in the model. The overall regression was statistically significant (R^2 = 0.21, F(5,72) = 3.82, p < .01). Results showed that having long hair (β = -0.26, p = .02) significantly predicted self-perceptions of conscientiousness.

Model 3 – Extraversion. Multiple linear regression was used to test if visual cues significantly predicted self-ratings of extraversion. Seventeen visual cues were included in the model. The overall regression was not statistically significant (R^2 =

0.35, $F(17,52) = 1.61$, $p = .10$) indicating that visual cues did not predict self-rated extraversion.

Model 4 – Agreeableness. Multiple linear regression was used to test if visual cues significantly predicted self-ratings of agreeableness. Eight visual cues were included in the model. The overall regression was statistically significant ($R^2 = 0.27$, $F(8,87) = 4.08$, $p < .01$). The results showed that wearing a necklace ($\beta = 0.24$, $p = .01$), dark light ($\beta = 0.27$, $p = .01$) and even light ($\beta = 0.29$, $p = .01$) significantly predicted self-perceptions of agreeableness.

Model 5 – Neuroticism. Multiple linear regression was used to test if visual cues significantly predicted self-ratings of neuroticism. Fifteen visual cues were included in the model. The overall regression was statistically significant ($R^2 = 0.53$, $F(15,53) = 3.97$, $p < .01$). The results showed that a vertical picture frame ($\beta = -0.2$, $p = .05$), dark light ($\beta = 0.25$, $p = .02$) and looking at the camera ($\beta = -0.35$, $p < .01$) significantly predicted self-perceptions of neuroticism.

ADEPT-15

Model 6 – Openness. Multiple linear regression was used to test if visual cues significantly predicted self-ratings of openness. Five visual cues were included in the model. The overall regression was statistically significant ($R^2 = 0.21$, $F(5,85) = 4.65$, $p < .01$). It was found that the presence of electronic equipment ($\beta = 0.25$, $p = .01$) significantly predicted self-ratings of openness.

Model 7 – Conscientiousness. Multiple linear regression was used to test if visual cues significantly predicted self-ratings of conscientiousness. Three visual cues were included in the model. The overall regression was statistically significant ($R^2 = 0.10$, $F(3,93) = 3.53$, $p = .02$), though none of the variables significantly contributed to the model.

Model 8 – Extraversion. Multiple linear regression was used to test if visual cues significantly predicted self-ratings of extraversion. Eight visual cues were included in the model. The overall regression was statistically significant ($R^2 = 0.25$, $F(8,79) = 3.28$, $p < .01$), though none of the variables significantly contributed to the model.

Model 9 – Agreeableness. Multiple linear regression was used to test if visual cues significantly predicted self-ratings of agreeableness. Twelve visual cues were included in the model. The overall regression was statistically significant ($R^2 = 0.39$, $F(12,56) = 3.03$, $p < .01$). The results showed that the presence of a coat rack ($\beta = 0.31$, $p < .01$) significantly predicted self-perceptions of agreeableness.

Model 10 – Neuroticism. Multiple linear regression was used to test if visual cues significantly predicted self-ratings of neuroticism. Two visual cues were included in the model. The overall regression was statistically significant ($R^2 = 0.10$, $F(2,94) = 4.99$, $p < .01$). It was found that a painting on the wall ($\beta = 0.24$, $p = .02$) and a poster/picture on the wall ($\beta = -0.23$, $p = .02$) both significantly predicted self-ratings of neuroticism.

Overall, results indicate that visual cues explain a significant amount of variance in self-reported personality scores. Specifically, the regression equations were found to be significant predictors of all but one personality trait – extraversion measured on the IPIP scale. This provides evidence of the first component in the Lens Model, cue validity, which is a requirement to demonstrate functional achievement. However, it must be noted that a substantial number of cues were not significant predictors within each model. The relative strength of each regression model is summarised for comparison in Table 23 (below). Due to the varying number of predictors used in each model, adjusted R^2 values are shown. As can be seen in the table, values for the ADEPT-15 models are generally lower those for the IPIP models. The next section will discuss results for the second component of the Lens Model, cue utilization.

Table 23. R squared and adjusted R-squared values in relation to cue **validity**.

Model		Traits	R^2	adjusted R^2
IPIP	1	Openness	.49	.42
	2	Conscientiousness	.21	.16
	3	Extraversion	.34	.13
	4	Agreeableness	.27	.21
	5	Neuroticism	.53	.40
ADEPT	6	Openness	.21	.17
	7	Conscientiousness	.10	.07
	8	Extraversion	.25	.17
	9	Agreeableness	.39	.26
	10	Neuroticism	.10	.08

Cue Utilization

The Lens Model purports that an observer uses environmental cues to infer a judgement about the underlying construct. Cue utilization refers to this link between cues (the lens) and observer inferences. In the current study, three sets of observer ratings are used. These are the Big 5 personality ratings gathered using three measurement methods: IPIP; ADEPT-15; and NLC-derived ratings.

As with the cue validity results described above, cue utilization will be evaluated using multiple linear regression with the visual cues as predictors and observer ratings of personality as the criterion. Visual cues were included in the regression equation where they significantly correlated with the personality trait predicted. Results of regression models using the Big 5 as measured by IPIP are reported first, followed by models using ADEPT-15 and then the NLC models. Again, all details for the models below can be found in Appendix XVI.

IPIP

Model 11 – Openness. Multiple linear regression was used to test if visual cues significantly predicted observer ratings of openness. Twelve visual cues were included in the model. The overall regression was statistically significant ($R^2 = 0.37$, $F(12,75) = 3.65$, $p < .001$). It was found that having only parts of the face showing ($\beta = -0.21$, $p = .03$), strong expressions ($\beta = 0.25$, $p = .01$) and pausing ($\beta = -0.26$, $p = .01$) significantly predicted perceptions of openness as rated by others.

Model 12 – Conscientiousness. Multiple linear regression was used to test if visual cues significantly predicted observer ratings of conscientiousness. Twenty visual cues were included in the model. The overall regression model was found to be statistically significant ($R^2 = 0.67$, $F(20,31) = 3.11$, $p = .002$). Results showed that colourful clothes ($\beta = 0.27$, $p = .049$), strong expressions ($\beta = 0.26$, $p = .04$) and looking at the camera ($\beta = 0.32$, $p = .04$) all significantly predicted others' ratings of conscientiousness.

Model 13 – Extraversion. Multiple linear regression was used to test if visual cues significantly predicted observer ratings of extraversion. Eleven visual cues were included in the model. The overall regression model was found to be statistically significant (R^2 = 0.39, $F(11,57)$ = 3.28, p = .002). Results showed that timid expressions (β = -0.31, p = .01) significantly predicted observer-rated extraversion.

Model 14 – Agreeableness. Multiple linear regression was used to test if visual cues significantly predicted observer ratings of agreeableness. Thirty visual cues were included in the model. The overall regression model was found to be statistically significant (R^2 = 0.91, $F(30,11)$ = 3.59, p = .015). It was found that a loose-fitting top (β = -0.77, p < .01), overexposed lighting (β = 0.54, p = .045), the presence of a pillow (β = -0.84, p = .01), rolling eyes (β = 0.82, p = .01), head oriented away from the camera (β = -0.75, p = .04), shoulder movement (β = -0.64, p = .02) and touching hair (β = 1.24, p = .02) significantly predicted ratings of agreeableness.

Model 15 – Neuroticism. Multiple linear regression was used to test if visual cues significantly predicted observer ratings of neuroticism. Twelve visual cues were included in the model. The overall regression model was statistically significant (R^2 = 0.43, $F(12,41)$ = 2.55, p = .013), however, none of the individual variables were found to be significant in the model.

ADEPT

Model 16 – Openness. Multiple linear regression was used to test if visual cues significantly predicted observer ratings of openness. Seventeen visual cues were included in the model. The overall regression was statistically significant (R^2 = 0.53, $F(17,46)$ = 3.05, p = .001). It was found that good personal cleanliness (β = 0.28, p = .049) and raised eyebrows (β = 0.28, p = .03) significantly predicted perceptions of openness as rated by others.

Model 17 – Conscientiousness. Multiple linear regression was used to test if visual cues significantly predicted observer ratings of conscientiousness. Twenty-four visual cues were included in the model. The overall regression was not statistically significant (R^2 = 0.65, $F(24,21)$ = 1.65, p = .125), showing that visual cues did not predict others' ratings of conscientiousness.

Model 18 – Extraversion. Multiple linear regression was used to test if visual cues significantly predicted observer ratings of extraversion. Twenty-seven visual cues were included in the model. The overall regression was not statistically significant (R^2 = 0.66, F(27,25) = 1.79, p = .074), indicating that visual cues did not predict observer-rated extraversion.

Model 19 – Agreeableness. Multiple linear regression was used to test if visual cues predicted observer ratings of agreeableness. Sixteen visual cues were included in the model. The overall regression was statistically significant (R^2 = 0.55, F(16,36) = 2.73, p = .006). Results indicated that good personal cleanliness (β = 0.38, p = .04) significantly predicted perceptions of agreeableness as rated by others.

Model 20 – Neuroticism. Multiple linear regression was used to test if visual cues predicted observer ratings of neuroticism. Twelve visual cues were included in the model. The overall regression was statistically significant (R^2 = 0.37, F(12,80) = 3.93, p < .001). Results showed that cheerful expressions (β = 0.29, p = .04), calm expressions (β = 0.25, p = .02) and arrogant-amused expressions (β = -0.34, p < .01) significantly predicted ratings of neuroticism by others.

NLC ADEPT

Model 21 – Openness. Multiple linear regression was used to test if visual cues predicted NLC derived ratings of openness. Nine visual cues were included in the model. The overall regression was not statistically significant (R^2 = 0.33, F(9,39) = 2.12, p = .052), indicating that the automatic ratings of openness were not predicted by visual cues.

Model 22 – Conscientiousness. Multiple linear regression was used to test if visual cues predicted automated NLC ratings of conscientiousness. Four visual cues were included in the model. The overall regression was statistically significant (R^2 = 0.18, F(4,74) = 4.17, p = .004). It was found that adjusting clothing (β = 0.26, p = .02) significantly predicted the automatic ratings of conscientiousness.

Model 23 – Extraversion. Multiple linear regression was used to test if visual cues predicted NLC derived ratings of extraversion. Nine visual cues were included in the model. The overall regression was statistically significant (R^2 = 0.37, F(9,51) =

3.30, $p = .003$). Results showed that a headset ($\beta = -0.36$, $p < .01$) and a completely visible face ($\beta = 0.35$, $p < .01$) significantly predicted NLC ratings of extraversion.

Model 24 – Agreeableness. Multiple linear regression was used to test if visual cues predicted NLC derived ratings of agreeableness. Nineteen visual cues were included as predictors in the model. The overall regression was statistically significant ($R^2 = 0.59$, $F(19,36) = 2.71$, $p = .005$), though none of the individual visual cues show significance in the model.

Model 25 – Neuroticism. Multiple linear regression was used to test if visual cues predicted automated NLC derived ratings of neuroticism. Twenty visual cues were included in the model. The overall regression was not significant ($R^2 = 0.52$, $F(20,21) = 1.12$, $p = .399$), suggesting that the automated ratings of neuroticism were not predicted by visual cues.

Overall, the regression results indicate that some of the variance in observer-rated personality traits is explained by visual cues. This provides evidence for cue utilization, suggesting that observers do make use of visual cues when making personality judgements about others. However, it must be noted that results across measurement methods (IPIP and ADEPT-15) were inconsistent with each other demonstrating limited overlap of cues. Furthermore, not all multiple regression models were significant. Similar to the cue validity results, the proportion of variance explained by visual cues tended to be higher for the trait measurements using IPIP compared to ADEPT-15. Importantly, the lowest explained variance was for the ADEPT NLC ratings. This is consistent with expectations as the algorithm works with text-based information and cannot make use of visual cues.

Taken together the cue validity and cue utilization results provide support for the potential functional achievement of the lenses. Functional achievement for each lens will be explored in the next section.

Table 24. R squared and adjusted R-squared values in rel to cue **utilization**.

Model		Traits	R^2	adjusted R^2
IPIP	11	Openness	.37	.27
	12	Conscientiousness	.67	.45
	13	Extraversion	.39	.27
	14	Agreeableness	.91	.65
	15	Neuroticism	.43	.26
ADEPT	16	Openness	.53	.36
	17	Conscientiousness	.65	.26
	18	Extraversion	.66	.29
	19	Agreeableness	.55	.35
	20	Neuroticism	.37	.28
ADEPT NLC	21	Openness	.33	.17
	22	Conscientiousness	.18	.14
	23	Extraversion	.37	.26
	24	Agreeableness	.59	.37
	25	Neuroticism	.52	.06

Functional Achievement

Functional achievement relates to the alignment between the initial two elements of the lens model, cue validity and cue utilization. Both cue utilization and cue validity are necessary but not sufficient to demonstrate functional achievement. Where a cue is valid, that is, it relates to the underlying construct, and a cue is utilised, that is, it relates to observer ratings of the construct, there is functional achievement. In this case, an observer is considered to have produced an accurate judgement about a construct through the lens of environmental cues. We can evaluate general functional achievement for each personality trait by examining the vector correlations between the cue validity side (self-ratings) and the cue utilization side (observer ratings). If cues are generally tending in the same direction for both cue validity and cue utilisation, it will result in a strong positive vector correlation. The vector correlation results for each trait are outlined in Table 25. Vector correlations for traits measured using IPIP are generally low and all are non-significant, ranging from -0.09 for openness to 0.16 for both conscientiousness and agreeableness. For ADEPT, three of the five traits show negative correlations with no vector correlation higher than 0.08 (neuroticism). These results suggest that despite the individual evidence for cue validity and cue utilization, there is little agreement between the two sides of the lenses. This can be further seen in examining the variables used for the respective cue validity and cue utilization regressions (see Appendix XVI) where there is little overlap between the cues. Overall, there is little evidence of functional achievement for visual cue lenses. Additional detailed findings are discussed below.

Table 25. Vector correlations of visual cues (self versus observer).

Model	Traits	Vector correlation
IPIP	Openness	-0.09
	Conscientiousness	0.16
	Extraversion	-0.03
	Agreeableness	0.16
	Neuroticism	-0.01
ADEPT	Openness	-0.26
	Conscientiousness	-0.27
	Extraversion	0.07
	Agreeableness	-0.20
	Neuroticism	0.08

Table 26. Regression analysis IPIP Openness.

	IPIP self-rating Openness					IPIP observer rating Openness				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	10.59	0	8.24 – 12.95	-0.17 – 0.17	**<0.001**	13.4	0	10.55 – 16.24	-0.18 – 0.18	**<0.001**
Pullover	0.29	0.13	-0.11 – 0.68	-0.05 – 0.31	0.154					
Hair cover face	0.59	0.29	0.21 – 0.97	0.10 – 0.47	**0.003**					
Earrings	1.03	0.31	0.45 – 1.62	0.13 – 0.49	**0.001**					
chair	-0.31	-0.13	-0.73 – 0.11	-0.31 – 0.05	0.146					
Map	1.17	0.21	0.17 – 2.18	0.03 – 0.39	**0.023**	0.49	0.1	-0.68 – 1.66	-0.14 – 0.34	0.405
Decoration	-0.41	-0.16	-0.92 – 0.10	-0.36 – 0.04	0.112					
Plant	-0.83	-0.15	-1.93 – 0.27	-0.34 – 0.05	0.136					
Heavy blinking	-0.87	-0.25	-1.67 – -0.07	-0.47 – -0.02	**0.033**					
Closed eyes	-0.64	-0.15	-1.67 – 0.39	-0.40 – 0.09	0.222					
Narrow eyes	-0.13	-0.04	-0.76 – 0.51	-0.24 – 0.16	0.692					
Pressed lips	-0.12	-0.04	-0.80 – 0.56	-0.24 – 0.17	0.73					
Business-like clothes						0.23	0.1	-0.19 – 0.64	-0.09 – 0.29	0.284
Only parts of face						-1.03	-0.21	-1.94 – -0.13	-0.40 – -0.03	**0.026**
Book shelve						0.12	0.05	-0.66 – 0.91	-0.26 – 0.36	0.753
Calendar						0.64	0.13	-0.45 – 1.74	-0.09 – 0.36	0.244
Book						0.26	0.13	-0.39 – 0.92	-0.19 – 0.45	0.428
Stationary						-0.1	-0.02	-1.00 – 0.80	-0.25 – 0.20	0.829
Athletic equipment						-0.78	-0.11	-2.16 – 0.61	-0.32 – 0.09	0.266
Friendly expressions						0.24	0.09	-0.32 – 0.80	-0.12 – 0.29	0.397
Timid expressions						-0.46	-0.12	-1.25 – 0.34	-0.33 – 0.09	0.259
Strong expressions						0.7	0.25	0.16 – 1.24	0.06 – 0.45	**0.012**
Pausing						-0.8	-0.26	-1.43 – -0.17	-0.46 – -0.05	**0.014**
Observations	84					88				
R2 / R2 adjusted	0.493 / 0.415					0.369 / 0.268				
F-statistic (*p*)	6.36 (**< 0.001**)					3.65 (**< 0.001**)				

Note. * p < 0.05, ** p < 0.01, *** p < 0.001.

Additional (detailed) findings

To explore the above findings in greater detail, correlations between self and observer ratings were examined. The full intercorrelation table is shown in Table 27. If observers were correctly inferring personality, a positive relationship between observer ratings of personality and self-ratings would be expected. The results show no significant relationships between self and observer ratings for the same trait using the IPIP or ADEPT measures, indicating that observers demonstrated little accuracy in personality judgements. In this context, it is understandable why functional achievement was not found. Intercorrelations within observer ratings were examined to further evaluate the rating process. For both IPIP and ADEPT-15, significant positive intercorrelations are seen across traits, suggesting that observers are not clearly differentiating between distinct personality traits. Specifically, for IPIP, ratings of openness, conscientiousness, extraversion and agreeableness are all positively correlated with values ranging from 0.22 (extraversion and agreeableness) to 0.60 (openness and conscientiousness). Similarly, for ADEPT-15, observer ratings of personality are positively correlated across all five traits with values ranging between 0.22 (extraversion and neuroticism) and 0.57 (agreeableness and neuroticism). The convergence of personality ratings suggests an inability by raters to differentiate between traits or it could indicate the presence of some underlying construct that is influencing all ratings.

Next, the relationship between ratings across the two measurement models was explored. Both personality measures attempt to measure the same construct – the Big 5 personality model. Therefore, a positive relationship is expected between ratings of the same trait across models. For self-ratings, ratings of openness are significantly negatively correlated (r = -0.32) with each other. In the case of observer ratings, results show significant positive correlations between ratings of conscientiousness (r = 0.32) and ratings of neuroticism (r = 0.20). All other traits show small, non-significant relationships. It should also be noted that, for the observer ratings, there were several significant positive relationships between different traits (for example, between IPIP ratings of openness and ADEPT ratings of conscientiousness), suggesting that the limited agreement that is present may be more due to the overall convergence of personality traits within scales than the consistency of ratings across scales.

Most problematic are the intercorrelations between observer ratings within the same method. For the ADEPT observer ratings this ranges from $r = 0.22$* between neuroticism and extraversion up to $r = 0.58$*** between openness and agreeableness. The same is true for the IPIP measurements, where the intercorrelation of observer ratings go up to $r = 0.60$*** between openness and conscientiousness. The observer ratings are a most crucial part for arguing functional achievements within the Lens Model. However, these high intercorrelations display a profound issue of how the observer rating has been conducted.

Not equally catastrophic, but showing the same trend are the intercorrelations of the self-ratings. Expecting low and no significant correlations, it is at least worth mentioning that multiple traits within both methods show significant correlations, such as between openness and conscientiousness ($r = 0.27$**) or between conscientiousness and extraversion ($r = 0.3$***) for IPIP self-ratings. As for the ADEPT method, the values are even higher such as between conscientiousness and neuroticism ($r = 0.45$***) or between extraversion and neuroticism ($r = 0.55$***).

The very high intercorrelations within the observer ratings could have led to putting the blame for not finding better functional achievement on the observers and their potentially lax ratings. It becomes clear that even the self-reports for both methods are somewhat questionable when looking at the respective intercorrelations.

While broad functional achievement was not found for any personality dimension, some individual cues do appear to show potential for functional achievement. For example, the *map* cue, related to the participant's environment and the *narrow eyes* cue, a dynamic cue in the face category, are in both the cue validity and cue utilization regressions for openness (see Appendix XVI Model 1 and Model 11). Specifically, the presence of a map was both positively related to self-perceptions of openness and observer ratings of openness while narrowing of the eyes was significantly negatively related to both ratings (see Table 28). More notable, however, is the lack of overlap of cues between the cue validity and cue utilization sides of the lens. Indeed, a similar pattern emerges in the remaining four dimensions (see Appendix XVII), with little-to-no evidence of functional achievement for specific cues and a high degree of non-overlap for each lens. Moreover, in the few cases where functional achievement is found, the result is not consistent across measurement methods. For example, for agreeableness, good personal cleanliness is positively related to both self and observer ratings on ADEPT-15, but not related to either self or observer ratings on IPIP. Overall, the general lack of alignment across measures together with the lack of overlap across each side of the lenses suggests that the limited functional achievement that is seen may be more coincidental than a true incidence of observer accuracy. In summary, the overall picture of the results can be comprised to the following:

(1) Having unexpected data patterns when it comes to the high intercorrelated observer ratings, as well as the too small intercorrelations between the used methods both for self and observer ratings.
(2) Very little general functional achievement, given the small vector correlation between self-ratings and observer ratings.
(3) Inconsistent patterns of individual visual cues, across traits, methods and sides of the lens.

With these diffuse findings, all results of Study 3 are to be handled with extreme caution. At this point, the clear recommendation is to not assume any external validity and to not infer any interpretations based on any of the findings from Study 3. Explicitly, this also means that the regression models and, with that, correlations for both sides of the lens have only limited validity. This point and its effects are further explored in Chapter 5.

Table 27. Intercorrelations for IPIP and ADEPT ratings.

		IPIP self-rating					ADEPT self-rating					IPIP observer rating					ADEPT observer rating			
		O	C	E	A	N	O	C	E	A	N	O	C	E	A	N	O	C	E	A
IPIP	C	0.27**																		
self	E	0.11	0.30**																	
rating	A	0.14	0.16	0.16																
	N	0.22*	0.16	0.18	0.04															
	O	-0.32**	-0.17	-0.14	-0.13	-0.03														
ADEPT	C	-0.10	-0.12	-0.04	-0.07	-0.11	0.28**													
self	E	-0.06	-0.17	0.02	-0.08	-0.00	0.16	0.24*												
rating	A	0.12	-0.14	-0.13	0.03	-0.12	0.02	0.18	0.1											
	N	-0.02	-0.15	0.04	-0.04	0.06	0.13	0.45***	0.55***	0.16										
IPIP	O	0.03	-0.02	0.17	-0.02	0.04	0.12	-0.11	-0.18	0.05	-0.09									
obs.	C	-0.01	-0.08	0.00	-0.07	-0.07	0.12	0.00	-0.03	0.24*	0.07	0.60***								
rating	E	-0.04	0.06	0.06	-0.03	0.09	0.08	-0.04	0.05	0.05	0.02	0.50***	0.47***							
	A	0.02	0.04	0.05	0.15	0.1	0.10	-0.04	-0.22*	0.14	-0.02	0.38***	0.42***	0.22*						
	N	0.05	0.13	0.01	0.15	-0.04	0.07	0.04	-0.07	0.06	-0.07	0.13	-0.06	-0.07	0.15					
ADEPT	O	-0.02	0.07	0.15	0.05	0.02	-0.06	-0.11	-0.08	-0.00	0.05	0.15	0.19	0.09	0.03	0.22*				
obs.	C	-0.10	0.03	0.01	0.00	-0.02	0.1	-0.02	-0.02	0.14	0.03	0.27**	0.32**	0.10	0.25*	0.03	0.40***			
rating	E	-0.04	0.17	-0.08	0.12	0.11	0.01	-0.16	-0.11	0.07	-0.08	0.01	0.05	0.12	-0.11	0.12	0.40***	0.27**		
	A	-0.09	0.08	-0.04	0.19	-0.08	-0.16	-0.21*	-0.23*	0.09	-0.24*	0.07	0.10	0.1	0.12	0.27**	0.58***	0.35***	0.38***	
	N	-0.18	-0.08	-0.06	0.05	-0.07	-0.08	-0.1	0.03	0.14	-0.07	0.10	0.21*	0.15	0.13	0.20*	0.32**	0.23*	0.22*	0.57***

Note. * p < 0.05, ** p < 0.01, *** p < 0.001.

Table 28. Lens model for Openness.

	cue validity		**Openness**	cue utilization		
	IPIP	ADEPT	Cue	IPIP	ADEPT obs.	ADEPT NLC
Appearance			Good personal cleanliness		0.28**	
			Wrinkled clothes			-0.25*
			Business-like clothes	0.23*		
		0.22*	Button-up shirt/blouse			
	.025*		Pullover			
			Loose-fitting top		-0.32**	
	0.22*		Hair cover face			
			Asymmetrical haircut			-0.25*
			Made-up eyes			0.23*
	0.23*		Earrings			-0.25*
			Tattoo		-0.22*	
			Headset		0.21*	
Media Properties			Only parts of face visible	-0.22*		
			High resolution image		0.20*	
Environment		-0.20*	Another pet present			
			Tidy room			0.25*
	0.23*		chair			
		0.20*	desk			
			Shelve			-0.21*
			Book shelve	0.20*		
			coat rack		0.21*	
			Book	0.20*		
	0.23*		**Map**	**0.21***		
			Stationary	0.21*		
			Bag		-0.23*	
		0.31**	Electronic equipment			
			Athletic equipment	-0.33**		
			Weapon		-0.21*	
	-0.20*		Decoration			
	-0.22*		Plant			
Face			Friendly expressions	0.22*	0.41***	
			Cheerful expressions		0.33**	
			Interested expressions		0.38***	
			Self-assured expressions		0.24*	
			Timid expressions	-0.29**		
			Indifferent expressions		-0.29**	
			Strong expressions	0.28**		
	-0.27**		Heavy blinking			
	-0.23*		Closed eyes			
	-0.21*		**Narrow eyes**			**-0.20***
			Raised eyebrows		0.22*	
			Smiling		0.35***	
	-0.13*		Pressed lips			
			Pausing	-0.28**		
			Fast mouth movements		0.27**	
Body			Maintained posture			-0.22*
			Open posture			0.28*
			Leaning forward		0.23*	
		0.21*	Head pulled back			
		0.22*	Sloped head posture			
	R^2	R^2		R^2	R^2	R^2
	0.49	0.26		0.41	0.55	0.30
	R^2 Adj.	R^2 Adj.		R^2 Adj.	R^2 Adj.	R^2 Adj.
	0.42	0.20		0.28	0.41	0.24

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Summary of results

This section is the final piece of all results of the three steps that have been taken throughout this body of work and links them to the research questions that have been introduced at the very beginning. For this, the research questions are recapped, and respective results are allocated to each of them in order to clarify whether and – given they have an exploratory nature – to what extent additional insights have been gathered per question.

Q1. Which visual cues can be captured during asynchronous video interviews?

This first research question has been answered in Step 1, where the definition, allocation and categorisation of the different visual cues have been processed. Based on that first step, Table 11 displays the preliminary number of cues (n = 236). However, after a respective data-cleaning process during Step 3 and in preparation for the final analyses, another 59 visual cues have been removed that tend to not differentiate enough or are too rare to be captured reliably. Therefore, the final list of visual cues that can be captured during asynchronous video interviews are displayed in Tables 16 to 20, split by category.

Q1.1. How can the captured visual cues be categorised and classified?

This question has also been answered in Step 1 and is likewise displayed in Table 11, namely by their content-related category (*Face, Body, Appearance, Media Properties* and *Environment*) as well as whether or not they are static and, therefore, unchanged during the video interview or dynamic and thereby show a different level of frequency throughout a video interview administration.

Q2. Which visual cues can be leveraged to predict observer ratings of different personality traits in asynchronous video interviews?

Table 29 (below) shows the visual cues that correlate significantly ($p < 0.05$, two-sided) with the different IPIP observer ratings. However, given the diffuse data pattern that was found in Study 3, the recommendation is to treat the findings of Study 3 lightly. In line with the stand that was taken in the previous chapter, the results most likely do not hold any external validity and, with that, should not be interpreted and generalised outside this body of work.

Table 29. visual cues that correlate (sig.) with IPIP observer ratings.

visual cue	**category**	**trait**
Business-like clothes	Appearances	Openness
Only parts of face visible (neg)	Media Properties	Openness
Book shelve	Environment	Openness
Book	Environment	Openness
Map	Environment	Openness
Stationary	Environment	Openness
Athletic equipment (neg)	Environment	Openness
Friendly expressions	Face	Openness
Timid expressions (neg)	Face	Openness
Strong expressions	Face	Openness
Pausing (neg)	Face	Openness
Colourful clothes	Appearances	Conscientiousness
Distinctive hair (neg)	Appearances	Conscientiousness
Good personal cleanliness	Appearances	Conscientiousness
Sitting	Body	Conscientiousness
Straight posture	Body	Conscientiousness
Athletic equipment (neg)	Environment	Conscientiousness
Bedroom (neg)	Environment	Conscientiousness
cookware & pots	Environment	Conscientiousness
Pillow (neg)	Environment	Conscientiousness
Shelve	Environment	Conscientiousness
Stationary	Environment	Conscientiousness
Toy	Environment	Conscientiousness
Looking at camera	Face	Conscientiousness
Looking down (neg)	Face	Conscientiousness
Pausing (neg)	Face	Conscientiousness
Strong expressions	Face	Conscientiousness
Timid expressions (neg)	Face	Conscientiousness
Only parts of face	Media Properties	Conscientiousness
Overexposed (neg)	Media Properties	Conscientiousness
Colourful clothes	Appearances	Extraversion
Long hair (neg)	Appearances	Extraversion
Pullover (neg)	Appearances	Extraversion
cookware & pots	Environment	Extraversion
Decoration	Environment	Extraversion
Window (neg)	Environment	Extraversion
Diverse facial expressions	Face	Extraversion
Rapidly changing facial expressions	Face	Extraversion
Strong expressions	Face	Extraversion
Timid expressions (neg)	Face	Extraversion

Tables continues on next page

Tables starts on previous page

Business-like clothes	Appearances	Agreeableness
Distinctive clothes (neg)	Appearances	Agreeableness
Distinctive hair (neg)	Appearances	Agreeableness
Loose-fitting top (neg)	Appearances	Agreeableness
Arms behind back (neg)	Body	Agreeableness
Body tilted (neg)	Body	Agreeableness
Body turned away (neg)	Body	Agreeableness
Empathetic gesture clapping (neg)	Body	Agreeableness
Head oriented away from camera (neg)	Body	Agreeableness
Leaning forward (neg)	Body	Agreeableness
Shoulder movement (neg)	Body	Agreeableness
Sloped head posture (neg)	Body	Agreeableness
Slouched posture (neg)	Body	Agreeableness
Straight posture (neg)	Body	Agreeableness
Touching body (neg)	Body	Agreeableness
Touching hair (neg)	Body	Agreeableness
Touching head (neg)	Body	Agreeableness
Biting lips (neg)	Face	Agreeableness
Duckface (neg)	Face	Agreeableness
Furrowed eyebrows (neg)	Face	Agreeableness
Pausing (neg)	Face	Agreeableness
Plopping lips (neg)	Face	Agreeableness
Pout (neg)	Face	Agreeableness
Pressed lips (neg)	Face	Agreeableness
Rolling eyes (neg)	Face	Agreeableness
Timid expressions (neg)	Face	Agreeableness
Big distance to wall (neg)	Media Properties	Agreeableness
Only parts of face (neg)	Media Properties	Agreeableness
Overexposed	Media Properties	Agreeableness
Asymmetrical haircut (neg)	Appearances	Neuroticism
Good personal cleanliness	Appearances	Neuroticism
Hair cover face (neg)	Appearances	Neuroticism
Heavy make up	Appearances	Neuroticism
Tended hair	Appearances	Neuroticism
Head tilt	Body	Neuroticism
Slouched posture	Body	Neuroticism
Kitchen	Environment	Neuroticism
Shelve (neg)	Environment	Neuroticism
Surprised expressions (neg)	Media Properties	Neuroticism
Light from right side (neg)	Media Properties	Neuroticism
Even light	Media Properties	Neuroticism

Nevertheless, an example will be given of how the results could be seen if the overall data pattern was more convincing. It must be clearly stated that this is purely an intellectual pastime, not a prompt or beginning of a practical guidance. The example will be made with the visual cues that link to the trait, Agreeableness.

To start with, of the 29 cues that correlate significantly with the observer rating, only two are non-negatively correlated. Most of the cues result from the categories Face and Body, and no cues from the category, Environment, have been found. Within the Appearances category, the candidates that seem to dress more formally, and have generally neat and non-distinctive appearances seemed to get higher ratings. When it comes to visual cues related to the category, Body, all are negatively correlated, hinting that the fewer body movements are displayed and the more stiff a candidate presents themselves, the higher on agreement they are rated by the observers. This is perfectly in line with face movements and expressions – the less mimical features are presented, the higher the Agreeable values. Lastly, there are only three Media property cues, which seem rather random and cannot easily be connected either within the category or to any of the other cues. None of them have been mentioned in any of the (non-academic) literature nor do they offer any obvious connection.

Q3. Which visual cues can be leveraged to predict self-ratings of different personality traits in asynchronous video interviews?

Similar to research question 2, the following table displays the visual cues that correlate with the respective self-ratings for the IPIP questionnaire (see Table 30). Again, as the data structure of Study 3 is not fully up to expectation, it is questionable to what degree the below findings are actually valid.

If they were, it is noticeable that a vast majority (<65%) of the below cues are negatively correlated with the self-ratings. In addition, those traits that seem to be linked to visual cues based on the preliminary literature findings, such as Openness and Extraversion, tend to show more significant visual cues in this study as well.

Table 30. visual cues that correlate (sig.) with IPIP self-ratings.

visual cue	**category**	**trait**
Pullover	Appearance	Openness
Hair cover face	Appearance	Openness
Earrings	Appearance	Openness
Chair	Environment	Openness
Map	Environment	Openness
Decoration (neg)	Environment	Openness
Plant (neg)	Environment	Openness
Heavy blinking (neg)	Face	Openness
Closed eyes (neg)	Face	Openness
Narrow eyes (neg)	Face	Openness
Pressed lips (neg)	Face	Openness
Long hair	Appearance	Conscientiousness
Nose piercing	Appearance	Conscientiousness
Empathetic gesture clapping (neg)	Body	Conscientiousness
Bag	Environment	Conscientiousness
chair (neg)	Environment	Conscientiousness
Laughing	Face	Conscientiousness
Long hair	Appearance	Extraversion
Shaking head (neg)	Body	Extraversion
Sloped head posture (neg)	Body	Extraversion
Athletic equipment (neg)	Environment	Extraversion
Wardrobe/closet (neg)	Environment	Extraversion
Window	Environment	Extraversion
Pausing (neg)	Face	Extraversion
Pout (neg)	Face	Extraversion
Raised eyebrows (neg)	Face	Extraversion
Rolling eyes (neg)	Face	Extraversion
Sceptical expressions (neg)	Face	Extraversion
Swallowing hard (neg)	Face	Extraversion
Wide open eyes (neg)	Face	Extraversion
Wrinkled forehead (neg)	Face	Extraversion
Camera level with head	Media Properties	Extraversion
Dark light	Media Properties	Extraversion
Overexposed (neg)	Media Properties	Extraversion
Necklace	Appearance	Agreeableness
Shelve (neg)	Environment	Agreeableness
Licking lips (neg)	Face	Agreeableness
Plopping lips (neg)	Face	Agreeableness
Surprised expressions (neg)	Face	Agreeableness
Wide open mouth (neg)	Face	Agreeableness
Dark light	Media Properties	Agreeableness
Even light	Media Properties	Agreeableness
Dyed hair	Appearance	Neuroticism
Shaking head (neg)	Body	Neuroticism
Calm expressions (neg)	Face	Neuroticism
Looking at camera (neg)	Face	Neuroticism
Raised eyebrows (neg)	Face	Neuroticism
Timid expressions (neg)	Face	Neuroticism
Wrinkled forehead (neg)	Face	Neuroticism
Dark light	Media Properties	Neuroticism
Face lit by screen (neg)	Media Properties	Neuroticism
Vertical picture frame (neg)	Media Properties	Neuroticism

Q4. Which visual cues can be leveraged to predict both self-ratings and observer ratings for the same personality traits in asynchronous video interviews?

As to be expected from previous result sections, the list for cues that correlate significantly with both self-ratings and observer ratings is rather short (see Table 31).

Table 31. visual cues that correlate (sig.) with both IPIP ratings.

visual cue	**category**	**trait**
Map	Environment	Openness
Long hair (neg)	Appearance	Extraversion
Window	Environment	Extraversion
Plopping lips	Face	Agreeableness

The low intercorrelations between IPIP and ADEPT ratings do not support the fact that these two methods can be used interchangeable for measuring the exact same constructs. Otherwise, there would have been more cues which correlate on the one side (i.e., self-ratings) with one instrument and on the other side (i.e., observer ratings) with the other instrument. If Table 27 had been closer to expectation, one could have argued differently. However, this does not leave any further wiggle room, which is why the above list of cues is the only one to be taken into account. For the same reason, the above tables only show cues related to IPIP and not to ADEPT. ADEPT was used to add additional data points and ease the link to include automatically-generated scores through the ADEPT NLC scoring. However, given the non-confirmatory results between IPIP and ADEPT, the focus remains on IPIP and pushes ADEPT further into the background, as – at least within this body of work – the data structure does not support that these two methods measure the same constructs.

Q5. Which visual cues can be leveraged to predict automatically-generated personality scores in asynchronous video interviews?

Lastly, in response to this research question, Table 32 shows visual cues that correlate significantly with the ADEPT NLC scores. Though in line with responses to Q3 and Q4, the suggestion is to not see those as valid predictors, given the problematic circumstances.

Table 32. visual cues that correlate (sig.) with ADEPT NLC ratings.

visual cue	**category**	**trait**
Wrinkled clothes (neg)	Appearances	Openness
Asymmetrical haircut (neg)	Appearances	Openness
Made-up eyes	Appearances	Openness
Earrings (neg)	Appearances	Openness
Tidy room	Environment	Openness
Shelve (neg)	Environment	Openness
Narrow eyes (neg)	Face	Openness
Maintained posture (neg)	Body	Openness
Open posture	Body	Openness
Headphones (neg)	Appearances	Conscientiousness
Adjust clothing	Body	Conscientiousness
Arrogant-amused expressions (neg)	Face	Conscientiousness
Vertical picture frame	Media Properties	Conscientiousness
Glasses	Appearances	Extraversion
Headset (neg)	Appearances	Extraversion
Long hair (neg)	Appearances	Extraversion
Standing	Body	Extraversion
Book shelve (neg)	Environment	Extraversion
Collection (neg)	Environment	Extraversion
Shelve (neg)	Environment	Extraversion
Toy (neg)	Environment	Extraversion
Complete face	Media Properties	Extraversion

Tables continues on next page

Tables starts on previous page

Clothing item (neg)	Appearances	Agreeableness
Good personal cleanliness	Appearances	Agreeableness
Highlights	Appearances	Agreeableness
Loose-fitting top (neg)	Appearances	Agreeableness
Necklace (neg)	Appearances	Agreeableness
Plucked eyebrows	Appearances	Agreeableness
Shaking legs (neg)	Body	Agreeableness
Bed (neg)	Environment	Agreeableness
Curtains (neg)	Environment	Agreeableness
Electronic equipment (neg)	Environment	Agreeableness
Pillow (neg)	Environment	Agreeableness
Stereo stand (neg)	Environment	Agreeableness
Friendly expressions	Face	Agreeableness
Heavy blinking (neg)	Face	Agreeableness
Pressed lips (neg)	Face	Agreeableness
Self-assured expressions	Face	Agreeableness
Timid expressions (neg)	Face	Agreeableness
High resolution image	Media Properties	Agreeableness
Overexposed	Media Properties	Agreeableness
Athletic clothes (neg)	Appearances	Neuroticism
Distinctive hair (neg)	Appearances	Neuroticism
Good personal cleanliness	Appearances	Neuroticism
Jewellery (neg)	Appearances	Neuroticism
Tattoo (neg)	Appearances	Neuroticism
Head oriented away from camera (neg)	Body	Neuroticism
Head pulled back (neg)	Body	Neuroticism
Maintained posture	Body	Neuroticism
Shaking head (neg)	Body	Neuroticism
Shoulder movement (neg)	Body	Neuroticism
Touching face (neg)	Body	Neuroticism
Electronic equipment (neg)	Environment	Neuroticism
Finished wall	Environment	Neuroticism
Stationary (neg)	Environment	Neuroticism
Stereo stand (neg)	Environment	Neuroticism
Wall pattern (neg)	Environment	Neuroticism
Weapon (neg)	Environment	Neuroticism
Duckface (neg)	Face	Neuroticism
Narrow eyes (neg)	Face	Neuroticism
Wide open eyes (neg)	Face	Neuroticism
Light halo (neg)	Media Properties	Neuroticism

Chapter 5: Conclusion

The goal of this body of research was first and foremost to determine which visual cues are detectable in asynchronous video interviews in selection processes and to then discover their relationship with the personality traits of the Five Factor Model. The discovery and clear display of this relationship is required for a second step, i.e., to enable research for the automatic and accurate prediction of personality traits of the video respondee using (solely) visual cues.

To get to the goal of this research initiative, a suited theoretical model has been selected, methods for generating the required dataset have been identified and respective studies have been conducted, alongside required analyses.

The purpose of this chapter is to review all findings, progress and insights that have been generated throughout this research initiative and set them into a broader context. To do so, first a short summary will be done to highlight the most relevant aspects of this work. Thereafter, a critical review is completed to investigate potential shortcomings that occurred during the course of this work. This is particularly relevant to the subsequent part which is to show implications to both research and applied practice.

Lastly, this chapter and with that this research initiative is closed with a final commentary and some personal notes.

Summary

The core of this research can be split into three empirical steps: (1) generating a visual cue list; (2) developing the visual cue inventory; and (3) generating and analysing data to answer the research questions – as well as the theoretical review that prefixes the first empirical step.

In detail, the theoretical review includes an evaluation and display of the components that are most relevant for the research that is to be completed. This includes components linked to video interviewing and visual cues, as well as the construct of personality. Therefore, relevant and recent literature and used models, such as the interview performance model, as well as definitions and explanations in the area of decision-making in employment interviews. In addition, the models and questionnaires used in the later studies to assess personality, ADEPT-15 and IPIP, are described as well as further methods and measurements used, such as visual cues and the Brunswik Lens Model.

Thereafter, the first empirical step contains all aspects to generate the visual cue list. A literature review was conducted, including 23 publications from relevant fields to generate 885 potential visual cues to be used. After applying pre-defined acceptance criteria that all visual cues would need to meet to be considered for this research, 172 visual cues were considered. In addition, a series of experiments has been launched to run two main Thinking-out-Loud studies to generate visual cues – specifically from the setting of asynchronous video interviews. The two combined studies generated an additional 167 visual cues after applying the same acceptance criteria. Further reduction of the visual cue list ended in a total of 236 visual cues that ultimately have been used in the final study.

In the second step, the visual cues needed to be formatted so that observers (coders) would be able to code whether or not each cue is present in a video. Hence, a visual cue inventory needed to be developed. The format and layout of the inventory needs to be defined. Furthermore, the exact process of how coders work with the inventory and how many visual cues they can process in each round need to be defined. For that purpose, in an adapted trial-and-error approach with two rounds and two different settings per round, the ideal number of visual cues that can be coded simultaneously is derived. Aligned with the outcome of the study, the visual cue inventory is split into three sections that are filled out individually to allow coders proper attention for all visual cues.

The last (third) step of the empirical process started with the creation of the database that contains self and observer personality ratings, as well as data from the fully-coded visual cue inventory. Thereafter, to answer the research questions and find out more about the relationship between visual cues and personality trait ratings, the respective analyses are conducted.

As a result, it turns out that regression models of visual cues were established to predict self-ratings of each personality trait, demonstrating cue validity. The same seemed to be true for cue utilization, where regression models of visual cues predicted respective observer ratings, again for both used methods (IPIP and ADEPT). However, running vector correlations between cues on the cue validity side and on the cue utilization side resulted in very low values, meaning there is little to no overlap of the value and direction between the cues that are used to predict self-ratings and those that are used to predict observer ratings.

This is a core finding and runs against the general direction of this research endeavour. Further examination of the dataset uncovered very high intercorrelations between traits within self-rating and observer ratings, but very little correlation between the same traits of the two methods used.

The conclusion for this work is that Study 3's results are to be treated very carefully. Even though individual relationships have been established between cues or groups of cues and personality traits, the general issue with the dataset does not allow a too wide use of those relationships.

Some of the relationships that have been uncovered are in line with the research findings and what was expected to be found. For instance, the cue 'good personal cleanliness' correlated with the observer rating of conscientiousness and, therefore, might be valid. However, most seem to be random and artifacts, which puts a general note of caution on the overall findings. I would feel very much at unease if any individual findings would be taken out of context and used or interpreted for further processing of automating visual cue coding for video interview settings.

Critical review

The findings of this research have the opportunity to add valuable insights into our understanding of visual cues and their role in asynchronous video interviews. They may impact and shape how visual cues in asynchronous video interviews can – or better: should *not* – be used.

Furthermore, this body of research is adding to the needed general research in this field that is relevant, given the progressive use of new technologies and, with that, its potential impact to selection processes (Chamorro-Premuzic et al., 2016).

It is to assume that our current understanding and ability to detect and interpret visual cues in asynchronous video interviews is limited. At this point, this body of work did not help to add to our understanding of linking self-rated personality traits, visual cues and observer ratings on the same traits. This suggests that visual cues from asynchronous video interviews should not yet be used for any fundamental conclusions or decision-making in practice and especially in high-stake situations, like a selection process.

However, this result has been driven mostly by the fact that the study findings in themselves are inconclusive. Therefore, there are several limitations that need to be pointed out to introduce improvement opportunities for similar or further research. Below are listed the major four issues that need to be tackled in a similar future research endeavour.

Issue 1: A narrow theory model approach

While this initiative was guided by the research carried out by Gosling and colleagues (2002) as stated in Chapter 2, it could still be questioned whether the amount of additional literature review comparing the used model, Brunswik's Lens Model, to other existing models of similar kind was too limited. Following the model and its pre-defined approach, the focus of this research may have been too narrow on identifying an extensive visual cue list. Efforts outside the targeted approach, such as investigating the impact of relationships between visual cues, have not been made.

For example, throughout the entire process of the research, the focus has been on identifying an extensive (ideally complete) list of visual cues, to then systematically code them and link them *individually* to different personality traits. At no point throughout the entire process has any effort been made to investigate the impact of relationships between visual cues. The defined cleaning rules helped to establish good analyses and – with that – reliable findings to the pre-defined approach. However, this approach does not shed light on the interaction of visual cues and potential moderative aspects of some. A factor analyses approach would also not have incorporated this issue to the full extent, as it would only approach it from a methodological angle. Rather, it would have required a different framework or addition to the linear relationship which is displayed between cues and traits in the described models. Not to undermine the current approach – most other models such as the Trait Reputation Identity Model by McAbee and Connelly (2016) or the Johari window by Luft and Ingham (1955), to name just two more, are in line with the ones discussed in this work. If a model would have been incorporated where moderation effects within cues or cues being linked to multiple traits were taken into account, some elements of the empirical part of this initiative could have been dealt with differently.

Issue 2: Shortcomings of the dataset and method

The second issue that potentially limits the applicability and generalisability of the findings is the shortcoming in diversity of the video database. The nature of working with videos is that they reveal significant personal information about the participants. That itself limits willingness to contribute in such studies. Additionally, generating the needed observation ratings and cue coding required additional participation which, in turn, put further pressure on the generation of video responses. All this resulted in leveraging the mTurk platform, which – among the respective advantages to cope for respective limitations – holds its own limitations. For instance, the majority of the observers, as well as the video respondees, are based in the United States and Europe. The dataset allows no cross-cultural application and even a generalisability within these regions is difficult as none of the analyses have been controlled for any demographic factors. This applies to other aspects of potential bias as well, examples are gender or race. It is well established that same gender between participants and observers influences the observer rating even for trained observers (Letzring, 2010; Booth et al., 2021). This could have had an impact in the present study and has not been controlled for.

Issue 3: Amount of data

Related to issue 2, but its own issue is the amount of available data. Simply put, it would have improved the situation – for reasons that will be discussed now – if more data would have been available. However, merely increasing the number of participants would not have changed any of the points from issue 2, which is why they are being discussed separately.

In addition to the reported results, factor analyses have been conducted to further explore and investigate the relationship of cues and their link to the measured traits (see Appendix XVIII). Similarly, further analyses have been done such as k-means clustering; however, not further reported in this body of work and only referenced in a related master thesis by Alnor (2021). One of the core methodological issues has been the ratio between visual cues (variables) and video respondees (cases). There are debates around the ideal variable-case ratio to allow factor analyses. The common ground in these discussions is that the number of cases should be higher than the number of variables. As for the present database, the number of variables (n = 236) exceeds the number of cases (n = 99).

Similarly, Gosling and colleagues (2005) recommend at least three coders per situation (in the present case: video). The database only consists of one coder per video, a shortcoming that does not allow to calculate intercoder agreement and could very likely affect reliability, a characteristic that is foundational for all assessments (Craik & Feimer, 1987). One could argue that capping the number of visual cues that are coded simultaneously as being defined in the design process is ensuring general high reliability. However, the inventory itself did not undergo a proper reliability study.

Issue 4: High intercorrelations

The intercorrelations of dataset 2 have been the basis for most of the withdrawing that has happened throughout this body of work, specifically for the results that have been generated in Study 3. Extensive rater training has been conducted prior to having gathered the observer ratings. In addition, all observers have been handpicked and trained by background, education and profession to be reliable observers in similar situations.

Likewise, to ensure proper participation, all video interviewees have been compensated more than for comparable studies/work on the mTurk platform. From general experience, this typically ensures high commitment and attracting participants that are typically more inclined to perform well.

Still, for both groups, the ratings have been problematic, at least from an intercorrelation perspective. Whether there is a more foundational issue with the questionnaires cannot be covered within this body of work. What remains is the conclusion that respective results – the visual cue correlations with respective self/observer ratings – cannot be fully trusted, thus rather treated as not valid.

Assets and Effort

These issues are not to fully undermine the results and outcome of this research. This body of research has been driven by the best of everyone's abilities. To put that into context, the time it took for the data collection alone is displayed in Table 33. It took around 28,000 minutes (or approximately 58 working days of eight hours each) of different participants to create the datasets, run the experiments and ultimately generate the visual cue-coding database with which the results of this research have been generated. This excludes all supporting time, transcripts of the Thinking-out-Loud experiments, rater-training or alike and only reflects the specific time it took to generate the datasets.

Moreover, in addition and supplementing this body of research, a total of two bachelor and two master theses have been conducted. A fifth thesis is on its way and will continue to shed light on the relationship between visual cues and personality traits. Not only did this initiative inspire students in their first steps of academic work, but also provided an opportunity for growth and development, an impact that is not to be underestimated for the benefit of the general field of research.

Table 33. Time it took to generate the available data.

area	type	N	Min per person	Total minutes
dataset 1	biographical data	142	5	710
dataset 1	ADEPT 15 self	142	20	2,840
dataset 1	6 video questions	142	24	3,408
pilot 1	Thinking out Loud	3	25	75
study 1	Thinking out Loud	10	30	300
pilot 1.5	Thinking out Loud	2	30	60
study 1.5	Thinking out Loud	10	30	300
study 2	systematic cue coding	10	15	150
dataset 2	biographical data	99	5	495
dataset 2	ADEPT 15 self	99	20	1,980
dataset 2	Mini-IPIP self	99	10	990
dataset 2	6 video questions	99	24	2,376
dataset 2	ADEPT 15 observer	99	35	3,465
dataset 2	mini-IPIP observer	99	35	3,465
dataset 2	systematic cue coding	99	75	7,425
Minutes spend to generate the data used in this research				**28,039**

Next steps for research and application

In this abstract, the focus will sit with different avenues that are provided with this research. To start with, a few specific next steps are outlined that can be continued within the spirit of this research endeavour. Next, a related but new approach is offered, still under the general idea of this body of research and interest. Lastly, a recommendation is given as which step – that might come natural specifically from an application perspective – should not (yet) follow.

Secondly, even though the research questions could not be answered to the extent it was possible for Gosling and colleagues (2005), there are practical implications that need to be addressed. One could argue that the findings are even more profound and relevant to be put into the context of practical usage to ensure highest scientific approach in practical applications.

Opportunities from a research perspective

The most prominent or obvious option to continue with this body of research is to enlarge the database that has already been created. This is also to tackle a specific issue that has been mentioned earlier in this work and would allow additional inferential analyses. The relevant upfront work is completed and all needed material, such as rater training documents, the inventory, as well as the respective Mini-IPIP questionnaires, are available so that a future researcher could start with the data collection right away. An additional database with the same methods, but different participants and observers would allow for a greater comparison and investigation of the issues found in this database. Are there, in fact, sample dependencies whatsoever related to observers or participants that drove the unexpected intercorrelations? Or are the same patterns found in an unrelated database as well? Either way, this would further add to the understanding of how the findings of this body of work can be interpreted and used.

This approach would, in theory, also help to drive additional opportunities, though they can be conducted on their own as well. Specifically, investigating more deeply into the relationship of self-rating, observer rating and NLC rating when controlled for various demographics, such as age, gender and level of education. Some of the issues related to the methods mentioned prior could be investigated this way. However, even more broadly, whether it be as part of a meta-analysis or completed with the dataset on its own, further investigations as to which degree similarities in demographics between target and observer are influencing personality judgements are needed. Given that this dataset includes an additional rating from the NLC-based scores, these analyses are of special interest to demonstrate its respective properties.

Taking a step further back, one could also pick up on the missing reliability characteristics of the visual cue inventory. If one were to look at the process of traditional test construction, there are specific characteristics that items need to be checked for, among others: item selectivity (Kline, 2016). Given the visual cue inventory acts as a traditional personality inventory and its items (cues) predict personality traits, it would be a good approach to apply similar quality criteria to its content. Item-writing criteria and characteristics such as difficulty, sensitivity, variance and intercorrelation should be incorporated and the visual cue inventory could be revised taking those aspects into consideration.

Lastly, and this is to take a step even further, one could take a very broad approach to using cues in asynchronous video interviews to predict personality. Non-verbal behaviours for instance are relevant to differentiate between individuals (Hall, Horgan & Murphy, 2019). When making (trait-based) inferences about one another, non-verbal behaviour is an essential influencing factor (Bonaccio et al., 2016). In that, non-verbal cues behave similarly to visual cues; however, they differ in not being in full control of the person, given it is difficult to suppress non-verbal cues (Roche & Arnold, 2018). A similar approach could be taken to investigate whether or not the findings of this study are replicable or different when targeting a different cue base. Ultimately, one could even combine visual and non-verbal cues for personality judgement, as they combine the aspects of cues that are present in interview settings but are essential for observer ratings in structured recruitment processes (Judge, Higgins & Cable, 2000) and therefore represent the blind spot of information that is deliberately excluded or simply unavailable during these processes.

Potentially different to most other research initiatives, at the end of this particular one, a word of caution is aired towards an intriguing but likely too early avenue of additional research. This research started with the problem statement that different sources of data is available during asynchronous video interviews but little insights yet to its best use under psychometric frameworks. This statement is still valid. The present work has not been able to shed additional light on data (cues) that should be incorporated for personality judgements. Therefore, at this stage, a clear recommendation is expressed to not proceed with investigations on how to automate the visual cue-coding process. It could have been a massive application advantage to automate the developed visual cue-coding inventory and its systematic cue-coding process, to then develop a scoring model that represents the regression models to predict personality traits. It is an intriguing thought, specifically when it comes to combine this approach with existing automatic-scoring methods such as the NLC-based one that was introduced in this work.

Even though a data-driven approach seems like a viable research avenue and the values derived in regression models from this work could be sufficient to further build on, relevant links of the Lens Model are missing as discussed when presenting the results earlier in this work. Taking individual findings and models out of the Lens Model context is not recommended.

Opportunities from a practitioner perspective

Before listing the various implications and learnings that can be taken from this work to the applied research field, one has to take into account the general limitation of that option. Taking the main study's results, the general tonality of taking this approach towards application is negative. It was not possible to answer the research questions that would allow direct practical application. However, there are a few aspects related to this work that are relevant to highlight and draw potential conclusions to their application in practice. Three areas are to be highlighted where this work can have practical influence.

Firstly, those insights that have been generated through the process allow for improvement as to how observer ratings are being conducted generally for asynchronous video interviews. Even though the situation and instructions are set so that ratings are completed and aligned with best practice (as outlined earlier in this work), it is nearly impossible to fully ignore snippets of visual cues (Ambady, Bernieri & Richeson, 2000; Ambady & Rosenthal, 1993). A good recommendation and an opportunity to further improve the rating process is to call out specific visual cues in the observer-training process. One would focus on visual cues that have very low cue validity in general as well as high correlation to multiple traits measured via observer ratings. Those visual cues could be specifically highlighted and mentioned during the training. Good examples based on the displayed (but potentially invalid) results would be the different expressions that candidates show during their recording, as well as their general gestures and body movements. Within the limited validity of the study, these seem not to have any cue validity but tend to have high cue utilization.

Refocusing observers back to the golden truth of interviews – what people actually say, not *how* they say it – and specifically calling out not to look too much for mimic and gestures might increase fair ratings and also further improve functional achievement as described by the Lens Model.

Another sensible way to deal with the nonexistence of a functional achievement but the existence of a cue utilization is to make this more prominent to participants recording the videos. Interviewees are already made aware of multiple aspects that they should factor in during the recording of videos, including proper lighting, position of the camera and similar media properties (Brenner, 2019). Highlighting the remaining visual cue categories and exemplary cues an observer might notice and take into account can help participants make a more-informed decision on how they want to present themselves – even including a higher risk of impression management.

Secondly, the results and, even further, the database that has been constructed can be leveraged to gain further insights to the construct validity of ADEPT. The development process and purpose has been pointed out in an earlier chapter. As a side effect to this body of work, it offers up a database to dive deeper into a comparison between the ADEPT model and a well-established five factor model. The sixth dimension of the ADEPT model has been neglected for most parts of the studies but has been captured both during self-ratings and observer ratings. The dataset therefore allows an in-depth comparison of the two models on self and observer level ratings to offer insights into the construct validity of ADEPT and to validate (or not) the ADEPT author's claim of the necessity of a sixth trait. This approach can be taken on in a secondary analysis using the dataset that has been generated in this body of research.

Thirdly, the dataset offers further insights into the validity of the NLC scoring of personality traits. Although only touched peripherally, the dataset allows further analyses to investigate the relationship between self-rating, observer rating and the role of the NLC rating as a third 'objective' rating. If one is to take the Johari window for instance by Funder (1995), the NLC rating allows a third dimension to enrich the model to a 3x3 grid. This perspective can offer additional insights into the interpretation of alignment, as well as misalignment between established ratings and the NLC rating. The dataset generated in this work, in turn, can be used to underpin this claim with data. One could have argued that this additional avenue might be one that is an opportunity for additional research rather than for practitioners. However, it is the author's strong belief that the further development of the displayed NLC rating has only a very niche application and thus, research interest might only apply to those in the immediate environment of AON. Therefore, further investigations in this direction will most likely be driven by AON, rather than independent researchers.

Final Commentary

At this last point of the write-up of my work, I allow myself the freedom to share a few personal comments with regards to the outcome.

Overall, I have been surprised and – at least at the beginning of working with the results – disappointed with the results of my work. I started this research eager to shed light on what seemed a promising opportunity to find more valid datapoints in modern psychometric tools, i.e., asynchronous video interviews. It took me a while to consider the lack of evidence for functional achievement in visual cues a reasonable and helpful outcome of this research.

For instance, different to the work led by Gosling, the section that is visible during the video – the literal window – is much more defined and limited compared to the more comprehensive view that Gosling took in his studies published in 2005. Most of the room (and with that the environmental cues) but similarly most of the body is typically not visible in asynchronous video interviews. The ratio of the cues visible versus the cues that are present but not visible is against those that are visible and therefore most likely extremely skewed. Secondly, the stakes of faking good and potentially higher (even in the manipulated mTurk study) and easier to do. Completely changing one's room is harder to 'only' adjust the angle of what is visible as well as adjusting the clothes and mimic of oneself.

However, finishing this work I am proud to be able to provide data points on the often-debated question in applied science whether or not visual cues can be used to predict personality traits. This research was a full step-wise, bottom-up approach, allowing to further tweak any of the steps and continue at whichever avenue to further add insights in this area. If at all possible, I would like to run the Study 3 for a second time with a complete new dataset, as this would most likely add the highest incremental value to the outcome of this body of work.

Disclosure

The data and core findings from Study 1 as well as the preceded literature review in Step 1 and the pilot study were leveraged and published in Maximilian Jansen's master thesis titled 'What you see is what you get: Exploring visual data in video interview', submitted 07.10.2019 at the Otto-Friedrich University Bamberg.

The data and core findings from Study 1.5 were leveraged and published in Anna Lüders' bachelor thesis titled 'Nutzbarkeit visueller Hinweisreize in asynchronen Videointerviews: eine Folgeuntersuchung', submitted 23.07.2020 at the Fachhochschule Westküste, Heide.

The data and some findings from Study 2 were leveraged and published in Svenja Hendrix's master thesis titled 'What Your Visuals Reveal About You: Assessing Personality by Visual Cues in Online Interviews', submitted 01.03.2021 at the Leuphana University Lüneburg.

Some of the data from dataset 2 were leveraged and published in Peter Alnor's master thesis titled 'Turning Disorder into Order - Clustering Visual Cues into Predictive Indicator Models as an Explorative Study with Video Interview Data', submitted 12.07.2021 at the Hochschule Fresenius, Hamburg.

Richard Justenhoven has been the main advisor to all four theses and has respectively contributed significantly to the designs, used methods and conclusions drawn. Due to university examination regulations at the Leuphana Universtiy Lüneburg that require advisors to hold the level of a Ph.D., Richard Justenhoven was not able to officially take the position of the co-advisor. Thankfully Dr. Achim Preuss, Richard Justenhoven's line manager at that time, acted on his behalf.

During the time of the work on this dissertation, Richard Justenhoven was employed by the Aon Assessment Solutions GmbH (Aon) as Director Global Products in Hamburg, Germany. Aon has had no influence into the design, data collection, conclusions or any other significant part of the presented work.

Eoin O'Callaghan is an employee of Aon Assessment Solutions and an expert in the usage of R. He has been supporting with the development of the R scripts that were used to generate some of the results for this dissertation. He has been reviewing, suggesting changes and test-running the various scripts.

Timothy Ilsley took on the job of reviewing and editing the fully-written manuscript. He is a skilled proofreader and linguistic expert – not an expert in psychology nor psychometrics. His recommendations have been most helpful and have been linguistic-based or to keep the central theme within the chapters. They have not been content related or contributed to the evolution of this work.

References

Abbasi, J. (2022). Pushed to Their Limits, 1 in 5 Physicians Intends to Leave Practice. JAMA, 327(15), 1435-1437.

Acikgoz, Y., Davison, K. H., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. International Journal of Selection and Assessment, 28(4), 399-416.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE access, 6, 52138-52160.

Agerström, J., & Rooth, D. O. (2011). The role of automatic obesity stereotypes in real hiring discrimination. Journal of Applied Psychology, 96(4), 790.

Allik, J., de Vries, R. E., & Realo, A. (2016). Why are moderators of self-other agreement difficult to establish?. Journal of Research in Personality, 63, 72-83.

Allik, J., Realo, A., Mõttus, R., & Kuppens, P. (2010). Generalizability of self-other agreement from one personality trait to another. Personality and Individual Differences, 48(2), 128–132.

Allport, G. W. (1937). Personality: A psychological interpretation.

Alnor, P. (2021). Turning Disorder into Order – Clustering Visual Cues into Predictive Indicator Models as an Explorative Study with Video Interview Data (Master's Thesis). Hochschule Fresenius, Hamburg, Germany.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. Journal of personality and social psychology, 64(3), 431.

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In M. P. Zanna (Ed.), Advances in experimental social psychology, 32, (pp. 201–271). Academic Press.

Ames, D. R., & Bianchi, E. C. (2008). The agreeableness asymmetry in first impressions: Perceivers' impulse to (mis) judge agreeableness and how it is moderated by power. Personality and Social Psychology Bulletin, 34(12), 1719-1736.

Anderson, B. (2021). 9 Tips for Conducting a Seamless Video Job Interview. Linkedin Talent Blog. https://www.linkedin.com/business/talent/blog/talent-acquisition/tips-for-conducting-seamless-virtual-job-interview. Retrieved 20.09.2022.

Andrews, L., Klein, S., Forseman, J., and Sachau, D. (2013). It's Easy Being Green: Benefits of Technology-Enabled Work. In Huffman, A. and Klein, S. (Eds). Green Organizations: Driving Change with I/O Psychology (pp. 149-169). Routledge, New York, USA.

Aon Assessment Solutions (2017a). vidAssess Technical Documentation. Hamburg, Germany.

Aon Assessment Solutions (2017b). ADEPT-15® Technical Documentation. Hamburg, Germany.

Apple Inc. (2007). Apple Reinvents the Phone with iPhone. Apple Newsroom Press Release. San Francisco, USA.

Apple Inc. (2010). Apple Launches iPad. Apple Newsroom Press Release. San Francisco, USA.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115.

Arseneault, R., & Roulin, N. (2021). A theoretical model of cross-cultural impression management in employment interviews. International Journal of Selection and Assessment, 29(3-4), 352-366.

Back, M. D., & Nestler, S. (2016). Accuracy of judging personality. In Hall, J. A., Mast, M. S., & West, T. V. (Eds.), The Social Psychology of Perceiving Others Accurately (pp. 98-124). Cambridge University Press.

Back, M. D., Schmukle, S. C., & Egloff, B. (2011). A Closer Look at First Sight: Social Relations Lens Model Analysis of Personality and Interpersonal Attraction at Zero Acquaintance. European Journal of Personality, 25(3), 225–238.

Back, M., Schmukle, S., & Egloff, B. (2008). How extraverted is honey.bunny77@hotmail.de?: Inferring personality from e-mail addresses. Journal of Research in Personality, 42(4), 1116–1122.

Baker, D. A., Burns, D. M., & Reynolds Kueny, C. (2020). Just sit Back and watch: large disparities between video and face-to-face interview observers in applicant ratings. International Journal of Human–Computer Interaction, 36(20), 1968-1979.

Bangerter, A., Roulin, N., & König, C. J. (2012). Personnel selection as a signaling game. Journal of Applied Psychology, 97(4), 719.

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.

Barrick, M. R., Shaffer, J. A., & DeGrassi, S. W. (2009). What you see may not be what you get: relationships among self-presentation tactics and ratings of interview and job performance. Journal of applied psychology, 94(6), 1394–1411.

Barrick, M. R., Swider, B. W., & Stewart, G. L. (2010). Initial Evaluations in the Interview: Relationships with Subsequent Interviewer Evaluations and Employment Offers. Journal of Applied Psychology, 95(6), 1163–1172. https://doi.org/10.1037/a0019918

Bartkoski, T., Lynch, E., Witt, C., & Rudolph, C. (2018). A meta-analysis of hiring discrimination against Muslims and Arabs. Personnel Assessment and Decisions, 4(2), 1.

Bartram, D. (2000). Internet recruitment and selection: Kissing frogs to find princes. International journal of selection and assessment, 8(4), 261-274.

Basch, J. M., & Melchers, K. G. (2019). Fair and flexible?! Explanations can improve applicant reactions toward asynchronous video interviews. Personnel Assessment and Decisions, 5(3), 2.

Basch, J. M., Brenner, F., Melchers, K. G., Krumm, S., Dräger, L., Herzer, H., & Schuwerk, E. (2021). A good thing takes time: The role of preparation time in asynchronous video interviews. International Journal of Selection and Assessment, 29(3-4), 378-392.

Basch, J. M., Melchers, K. G., Kurz, A., Krieger, M., & Miller, L. (2021). It takes more than a good camera: which factors contribute to differences between face-to-face interviews and videoconference interviews regarding performance ratings and interviewee perceptions?. Journal of business and psychology, 36(5), 921-940.

Batrinca, L. M., Mana, N., Lepri, B., Pianesi, F., & Sebe, N. (2011). Please, tell me about yourself: automatic personality assessment using short self-presentations. In Proceedings of the 13th international conference on multimodal interfaces (pp. 255-262). ACM.

Bauer, T. N., Truxillo, D. M., Paronto, M. E., Weekley, J. A., & Campion, M. A. (2004). Applicant reactions to different selection technology: Face-to-face, interactive voice response, and computer-assisted telephone screening interviews. International Journal of Selection and Assessment, 12(1-2), 135-148.

Beer, A., & Brooks, C. (2011). Information quality in personality judgment: The value of personal disclosure. Journal of Research in Personality, 45(2), 175-185.

Behrend, T. S., & Thompson, L. F. (2013). Combining I-O psychology and technology for an environmentally sustainable world. In A. H. Huffman & S. R. Klein (Eds.), Green organizations: Driving change with I-O psychology (pp. 300–322). Routledge, New York, USA

Behrend, T., Toaddy, S., Thompson, L. F., & Sharek, D. J. (2012). The effects of avatar appearance on interviewer ratings in virtual employment interviews. Computers in Human Behavior, 28(6), 2128-2133.

Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. Psychological review, 81(6), 506.

Beukeboom, C. J., Tanis, M., & Vermeulen, I. E. (2013). The language of extraversion: Extraverted people talk more abstractly, introverts are more concrete. Journal of Language and Social Psychology, 32(2), 191–201.

Blacksmith, N., & Poeppelman, T. (2014). Video-based technology: The next generation of recruitment and hiring. TIP: The Industrial-Organizational Psychologist, 52(2), 84-88.

Blacksmith, N., Willford, J. C., & Behrend, T. S. (2016). Technology in the employment interview: A Meta-Analysis and Future Research Agenda. Personnel Assessment and Decisions, 2(1), 2.

Bonaccio, S., O'Reilly, J., O'Sullivan, S. L., & Chiocchio, F. (2016). Nonverbal behavior and communication in the workplace: A review and an agenda for research. Journal of Management, 42(5), 1044-1074.

Booth, B. M., Hickman, L., Subburaj, S. K., Tay, L., Woo, S. E., D'Mello, S. K. (2021). Bias and Fairness in Multimodal Machine Learning: A Case Study of Automated Video Interviews. In Proceedings of the 2021 International Conference on Multimodal Interaction, Montréal, Canada.

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. Journal of personality and social psychology, 62(4), 645.

Borkenau, P., & Liebler, A. (1995). Observable Attributes as Manifestations and Cues of Personality and Intelligence. Journal of Personality, 63(1), 1-25.

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. Journal of personality and social psychology, 86(4), 599.

Borsellino, R. (n. d.). 20 Video Interview Tips to Help You Dazzle the Hiring Manager and Get the Job. The Muse. https://www.themuse.com/advice/video-interview-tips. Retrieved 09.09.2022.

Bourdage, J. S., Roulin, N., & Tarraf, R. (2018). "I (might be) just that good": Honest and deceptive impression management in employment interviews. Personnel Psychology, (July 2016), 1–36.

Bowles, M. A. (2010). The think-aloud controversy in second language research. Routledge, New York, USA.

Brandt, O., Justenhoven, R. T., Schöffel, M., (2020). Web-basierte Videointerviews. In K. P. Stulle (Ed). Digitalisierung der Management-Diagnostik (pp. 43-66). Springer, Köln, Germany.

Breil, S. M., Osterholz, S., Nestler, S., & Back, M. D. (2021). Contributions of Nonverbal Cues to the Accurate Judgment of Personality Traits. In T.D. Letzring, & J. S. Spain (Eds.). The Oxford handbook of accurate personality judgment (pp. 195-218). Oxford University Press, Oxford, United Kingdom.

Brenner, F. S. (2019). Asynchronous Video Interviews in Selection: A Systematic Review and Five Empirical Investigations (Doctoral dissertation). Freie Universität Berlin, Berlin, Germany.

Brenner, F. S., Ortner, T. M., & Fay, D. (2016). Asynchronous video interviewing as a new technology in personnel selection: The applicant's point of view. Frontiers in Psychology, 7, 863-874.

Brunswik, E. (1956). Perception and the representative design of psychological experiments. Berkeley: University of California Press.

Burkov, A. (2019). The Hundred-Page Machine Learning Book. Quebec, Canada.

Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. Journal of Behavioral Decision Making, 33(2), 220-239.

Cammio, (2021). Our Platform. https://cammio.com/our-platform/

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. Personnel psychology, 50(3), 655-702.

Campion, M. C., Campion, M. A., Campion, E. D., & Reider, M. H. (2016). Initial investigation into computer scoring of candidate essays for personnel selection. Journal of Applied Psychology, 101(7), 958.

Cannata, D., Breil, S. M., Back, M., Lepri, B., & O'Hora, D. (in press). Toward an Integrative Approach to Nonverbal Personality Detection: Connecting

Psychological and Artificial Intelligence Research. Technology, Mind and Behavior.

Cantador, I., Fernández-Tobías, I., & Bellogín, A. (2013). Relating Personality Types with User Preferences in Multiple Entertainment Domains. UMAP Workshops.

Carlson, J. R., & Zmud, R. W. (1999). Channel expansion theory and the experiential nature of media richness perceptions. Academy of management journal, 42(2), 153-170.

Chamorro-Premuzic, T., Akhtar, R., Winsborough, D., & Sherman, R. A. (2017). The datafication of talent: How technology is advancing the science of human potential at work. Current Opinion in Behavioral Sciences, 18, 13-16.

Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New talent signals: Shiny new objects or a brave new world? Industrial and Organizational Psychology-Perspectives on Science and Practice, 9(3), 621-640.

Chapman, D. S., & Rowe, P. M. (2001). The impact of videoconference technology, interview structure, and interviewer gender on interviewer evaluations in the employment interview: A field experiment. Journal of Occupational and Organizational Psychology, 74(3), 279-298.

Chapman, D. S., & Rowe, P. M. (2002). The influence of videoconference technology and interview structure on the recruiting function of the employment interview: A field experiment. International Journal of Selection and Assessment, 10(3), 185-197.

Chapman, D. S., & Zweig, D. I. (2005). Developing a nomological network for interview structure: Antecedents and consequences of the structured selection interview. Personnel Psychology, 58(3), 673-702.

Chapman, D. S., Uggerslev, K. L., & Webster, J. (2003). Applicant reactions to face-to-face and technology-mediated interviews: A field investigation. Journal of Applied Psychology, 88(5), 944-953.

Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. Brock Education: A Journal of Educational Research and Practice, 12(2), 68-82.

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. Human Performance, 18, 267-307.

Civil Rights Act, 42 USCS § 2000e (1964).

Cohen, A. (2021). How to Quit Your Job in the Great Post-Pandemic Resignation Boom. Bloomberg Businessweek. https://www.bloomberg.com/news/articles/2021-05-10/quit-your-job-how-to-resign-after-covid-pandemic?cmpid=socialflow-twitter-businessweek&utm_medium=social&utm_content=businessweek&utm_source=twitter&utm_campaign=socialflow-organic#xj4y7vzkg. Retrieved 20.07.2022.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46.

Colvin, C. R. (1993). " Judgable" people: Personality, behavior, and competing explanations. Journal of Personality and Social Psychology, 64(5), 861.

Colvin, C. R., & Funder, D. C. (1991). Predicting personality and behavior: a boundary on the acquaintanceship effect. Journal of Personality and Social Psychology, 60(6), 884.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. Journal of management, 37(1), 39-67.

Connelly, B. S., McAbee, S. T., Oh, I. S., Jung, Y., & Jung, C. W. (2021). A multirater perspective on personality and performance: An empirical examination of the trait–reputation–identity model. Journal of Applied Psychology.

Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. International Journal of Selection and Assessment, 15(1), 110-117.

Converse, P. D., Peterson, M. H., & Griffith, R.L. (2009). Faking on personality measures: Implications for selection involving multiple predictors. International Journal of Selection and Assessment, 17, 47-60.

CooperLomaz (2020). Video Interviews: how to showcase your personality. CooperLomaz Blog. https://www.cooperlomaz.co.uk/blog/view/192/Video-Interviews-How-To-Showcase-Your-Personality. Retrieved 10.09.2020.

Costa, P. T., & McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO-PI-R) professional manual. Psychological Assessment Resources, Odessa, USA.

Costa, P. T., Jr., & McCrae, R. R. (1985). The NEO Personality Inventory manual. Psychological Assessment Resources, Odessa, USA.

Costa, P. T., Jr., & McCrae, R. R. (1991). Adding liebe und arbeit: The full five-factor model and well-being. Personality and Social Psychology Bulletin, 17(2), 227-232.

Costa, P. T., Jr., & McCrae, R. R. (2005). The NEO-PI-3: a more readable revised NEO Personality Inventory. Journal of Personality Assessments, 84(3), 261-270.

Council Directive 2000/78/EC establishing a general framework for equal treatment in employment and occupation (2000). Official Journal L 303, 02/12/2000 P. 0016 – 0022.

Craik, K. H., & Feimer, N. R. (1987). Environmental assessment. In D. Stokols & I. Altman (Eds.), Handbook of environmental psychology (Vol. 2, pp. 891-917). John Wiley, New York, USA.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. Management Science, 32(5), 554-571. https://doi.org/10.1287/mnsc.32.5.554.

Dattner, B., Chamorro-Premuzic, T., Buchband, R., & Schettler, L. (2019). The legal and ethical implications of using AI in hiring. Harvard Business Review, 25.

De Kock, F. S., Lievens, F., & Born, M. P. (2020). The profile of the 'Good Judge'in HRM: A systematic review and agenda for future research. Human Resource Management Review, 30(2), 100667.

de Villiers, C., Farooq, M. B., & Molinari, M. (2021). Qualitative research interviews using online video technology–challenges and opportunities. Meditari Accountancy Research.

Deakin, H., & Wakefield, K. (2014). Skype interviewing: Reflections of two PhD researchers. Qualitative research, 14(5), 603-616.

DeCarlo, L. (n.d.). 12 Tips for One-Way Video Interview Success (Prerecorded Interviews). Job Hunt. https://www.job-hunt.org/handling-one-way-video-interviews/. Retrieved 10.09.2022.

DeGroot, T., & Gooty, J. (2009). Can nonverbal cues be used to make meaningful personality attributions in employment interviews?. Journal of business and psychology, 24(2), 179-192.

DeGroot, T., & Kluemper, D. (2007). Evidence of predictive and incremental validity of personality factors, vocal attractiveness and the situational interview. International Journal of Selection and Assessment, 15(1), 30-39.

DeGroot, T., & Motowidlo, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. Journal of Applied Psychology, 84(6), 986.

DeLeon, M. (2015). What Really Happens When You Hire The Wrong Candidate. Entrepreneur. https://www.entrepreneur.com/article/244730. Retrieved 28.06.2022.

Dennis, A. R., & Valacich, J. S. (1999). Rethinking media richness: Towards a theory of media synchronicity. In Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers (pp. 10-pp). IEEE.

Dennis, A. R., Fuller, R. M., & Valacich, J. S. (2008). Media, tasks, and communication processes: A theory of media synchronicity. MIS quarterly, 575-600.

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. Psychological Bulletin, 111(2), 203–243.

Derous, E., Buijsrogge, A., Roulin, N., & Duyck, W. (2016). Why your stigma isn't hired: A dual-process framework of interview bias. Human Resource Management Review, 26(2), 90-111.

DeYoung, C. G., Quilty, L. C., & Peterson, J.B. (2007). Between facets and domains: 10 aspects of the Big Five. Journal of Personality and Social Psychology, 93(5), 880-896.

Di Domenico, S. I., Quitasol, M. N., & Fournier, M. A. (2015). Ratings of conscientiousness from physical appearance predict undergraduate academic performance. Journal of Nonverbal Behavior, 39(4), 339-353.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114.

Drasgow, F., Chernyshenko, O. S., & Stark, S. (2009). Test theory and personality measurement. In J. N. Butcher (Ed.), Oxford handbook of personality assessment (pp. 59–80). Oxford University Press, Oxford, United Kingdom.

Dunlop, P. D., Holtrop, D., & Wee, S. (2022). How asynchronous video interviews are used in practice: A study of an Australian-based AVI vendor. International Journal of Selection and Assessment.

Eisenkraft, N. (2013). Accurate by way of aggregation: Should you trust your intuition-based first impressions?. Journal of Experimental Social Psychology, 49(2), 277-279.

Eppler, M. J., & Kernbach, S. (2016). Dynagrams: Enhancing Design Thinking Through Dynamic Diagrams. In W. Brenner & F. Uebernickel (Eds.), Design Thinking for Innovation (pp. 85-102). Springer, Cham, Switzerland.

Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. The Cambridge handbook of expertise and expert performance, 223-241.

European Commission (2019). Ethics guidelines for trustworthy AI. High-Level Expert Group on Artificial Intelligence. Brussels, Belgium.

Florea, L., Valcea, S., Hamdani, M. R., & Dougherty, T. W. (2018). From first impressions to selection decisions: The role of dispositional cognitive motivations in the employment interview. Personnel review.

Forgas, J. P. (2011). Can negative affect eliminate the power of first impressions? Affective influences on primacy and recency effects in impression formation. Journal of Experimental Social Psychology, 47(2), 425-429.

Frauendorfer, D., & Mast, M. S. (2015). The impact of nonverbal behavior in the job interview. In The social psychology of nonverbal communication (pp. 220-247). Palgrave Macmillan, London.

Frieder, R. E., Van Iddekinge, C. H., & Raymark, P. H. (2016). How quickly do interviewers reach decisions? An examination of interviewers' decision-making time across applicants. Journal of Occupational and Organizational Psychology, 89(2), 223-248.

Funder, D. C. (1995). On the accuracy of personality judgment: a realistic approach. Psychological review, 102(4), 652-670.

Funder, D. C. (2012). Accurate personality judgment. Current Directions in Psychological Science, 21(3), 177-182.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: acquaintanceship, agreement, and the accuracy of personality judgment. Journal of personality and social psychology, 55(1), 149.

Gill, A. J., & Oberlander, J. (2003). Perception of e-mail personality at zero-acquaintance: Extraversion takes care of itself; neuroticism is a worry. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 25, No. 25).

Gliozzo, A., Ackerson, C., Bhattacharya, R., Goering, A., Jumba, A., Kim, S. Y., Krishnamurthy, L., Lam, T., Littera, A., McIntosh, I., Murthy, S., & Ribas, M. (2017). Building Cognitive Applications with IBM Watson Services: Volume 1 – Getting started. IBM Redbooks.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I.

Deary, F. De Fruyt, & F. Ostendorf (Eds.), Personality Psychology in Europe, Vol. 7 (pp. 7-28). Tilburg University Press, Tilburg, The Netherlands.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. Journal of Research in Personality, 40, 84-96.

Golden, R. (2020). COVID-19 has made hiring more efficient, report says. HR Dive Dive Brief. https://www.hrdive.com/news/covid-19-has-made-hiring-more-efficient-report-says/587059/. Retrieved 27.09.2022.

Gonzalez, M. F., Capman, J. F., Oswald, F. L., Theys, E. R., & Tomczak, D. L. (2019). “Where’s the IO?” Artificial intelligence and machine learning in talent management systems. Personnel Assessment and Decisions, 5(3), 5.

Gonzalez, M. F., Liu, W., Shirase, L., Tomczak, D. L., Lobbe, C. E., Justenhoven, R. T., & Martin, N. R. (2022). Allying with AI? Reactions Toward Human-Based, AI/ML-Based, and Augmented Hiring Processes. Computers in Human Behavior, 130.

Gorden, R. L. (1998). Basic Interviewing Skills. Waveland Press, Illinois, USA.

Gorman, C. A., Robinson, J., & Gamble, J. S. (2018). An investigation into the validity of asynchronous web-based video employment-interview ratings. Consulting Psychology Journal: Practice and Research, 70(2), 129.

Gosling, S. D., Craik, K. H., Martin, N. R., & Pryor, M. R. (2005). The personal living space cue inventory: An analysis and evaluation. Environment and Behavior, 37(5), 683-705.

Gosling, S. D., Gifford, R., & McCunn, L. J. (2013). The selection, creation, and perception of interior spaces: An environmental psychology approach. In G. Booker, & L. Weinthal (Eds.), Handbook of interior architecture and design (pp. 278–290). Berg.

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: personality judgments based on offices and bedrooms. Journal of personality and social psychology, 82(3), 379.

Graham, L. T., Gosling, S. D., & Travis, C. K. (2015). The psychology of home environments: A call for research on residential space. Perspectives on Psychological Science, 10(3), 346-356.

Graham, L. T., Sandy, C. J., & Gosling, S. D. (2011). Manifestations of individual differences in physical and virtual environments. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), The Wiley-Blackwell handbook of individual differences (pp. 773–800). Wiley Blackwell, New Jersey, USA.

Graham, L., & Gosling, S. D. (2021). Can the Ambiance of a Place be Determined by the User Profiles of the People Who Visit It?. Proceedings of the International AAAI Conference on Web and Social Media, 5(1), 145-152.

Griswold, K. R., Phillips, J. M., Kim, M. S., Mondragon, N., Liff, J., & Gully, S. M. (2021). Global differences in applicant reactions to virtual interview synchronicity. The International Journal of Human Resource Management, 1-28.

Hall, J. A., Back, M. D., Nestler, S., Frauendorfer, D., Schmid Mast, M., & Ruben, M. A. (2018). How do different ways of measuring individual differences in zero-acquaintance personality judgment accuracy correlate with each other?. Journal of Personality, 86(2), 220-232.

Hall, J. A., Horgan, T. G., & Murphy, N. A. (2019). Nonverbal communication. Annual Review of Psychology, 70(1), 271–294.

Hall, J. A., Mast, M. S., & West, T. V. (2016). Accurate interpersonal perception: Many traditions, one topic. In Hall, J. A., Mast, M. S., & West, T. V. (Eds.), The Social Psychology of Perceiving Others Accurately (pp. 3-22). Cambridge University Press.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Sage, USA.

Hendrix, S. (2021). What your Visuals Reveal About You: Assessing Personality by Visual Cues in Online Interviews (Master's Thesis). Leuphana Universität, Lüneburg, Germany.

Hickman, L., Bosch, N., Ng, V., Saef, R., Tay, L., & Woo, S. E. (2021). Automated video interview personality assessments: Reliability, validity, and generalizability investigations. Journal of Applied Psychology.

Hiemstra, A. M., Oostrom, J. K., Derous, E., Serlie, A. W., & Born, M. P. (2019). Applicant perceptions of initial job candidate screening with asynchronous job interviews: Does personality matter?. Journal of Personnel Psychology, 18(3), 138.

Hirschmüller, S., Egloff, B., Nestler, S., & Back, M. D. (2013). The dual lens model: A comprehensive framework for understanding self–other agreement of personality judgments at zero acquaintance. Journal of Personality and Social Psychology, 104(2), 335.

Huffcutt, A. I. (2011). An empirical review of the employment interview construct literature. International Journal of Selection and Assessment, 19(1), 62-81.

Huffcutt, A. I., & Youngcourt, S. S. (2007). Employment interviews. In D. L. Whetzel & G. R Wheaton (Eds.), Applied measurement: Industrial psychology in human resource management (pp. 181–199). Psychology Press.

Huffcutt, A. I., Van Iddekinge, C. H., & Roth, P. L. (2011). Understanding applicant behavior in employment interviews: A theoretical model of interviewee performance. Human Resource Management Review, 21(4), 353-367

Human, L. J., & Biesanz, J. C. (2011a). Target adjustment and self-other agreement: Utilizing trait observability to disentangle judgeability and self-knowledge. Journal of Personality and Social Psychology, 101(1), 202.

Human, L. J., & Biesanz, J. C. (2011b). Accuracy and assumed similarity in first impressions of personality: Differing associations at different levels of analysis. Journal of Research in Personality, 46(1), 106-110.

Human, L. J., & Biesanz, J. C. (2013). Targeting the good target an integrative review of the characteristics and consequences of being accurately perceived. Personality and Social Psychology Review, 17, 248 –272

Human, L. J., Biesanz, J. C., Finseth, S. M., Pierce, B., & Le, M. (2014). To thine own self be true: Psychological adjustment promotes judgeability via personality–behavior congruence. Journal of Personality and Social Psychology, 106(2), 286.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. Psychological bulletin, 96(1), 72.

Indurkhya, N., & Damerau, F. J. (2nd Ed.). (2010). Handbook of natural language processing. Routledge, Oxford, United Kingdom.

Jansen, M. R. (2019). What you see is what you get: Exploring visual data in video interviews (Master’s Thesis). Universität Bamberg, Bamberg, Germany.

Jansen, M. R., Justenhoven, R. T., Hock, M., & Krumm, S. (2020). What you see is what you get: Exploring visual data in video interviews. Poster at the GWPS Conference, Stuttgart, Germany.

Javed, A., & Brishti, J. K. (2020). The viability of AI-based recruitment: A systematic literature review (Magister’s Thesis). Umea University, Sweden.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. Journal of personality, 61(4), 521-551.

Joshi, A., Bloom, D. A., Spencer, A., Gaetke-Udager, K., & Cohan, R. H. (2020). Video interviewing: a review and recommendations for implementation in the era of COVID-19 and beyond. Academic radiology, 27(9), 1316-1322.

Kanning, U. P. (2018). Standards der Personaldiagnostik. Hogrefe Verlag.

Kanning, U. P., Kraul, L. F., & Litz, R. Z. (2019). Einstellungen zu digitalen Methoden der Personalauswahl. Journal of Business and Media Psychology, 10, 57-61.

Kasmar, J. V. (1970). The Development of a Usable Lexicon of Environmental Descriptors. Environment and Behavior, 2(2), 247.

Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. Psychological Review, 98, 155–163.

Kline, P. (2016). A Handbook of Test Construction: Introduction to Psychometric Design. Routledge, Oxford, United Kingdom.

Klupacs, L. (n.d.). 9 ways to perfect your video interview. Accenture Careers Blog. https://www.accenture.com/us-en/blogs/blogs-careers/9-ways-to-perfect-your-video-interview. Retrieved 10.09.2022.

Köchling, A., Riazy, S., Wehner, M. C., & Simbeck, K. (2021). Highly accurate, but still discriminatory. Business & Information Systems Engineering, 63(1), 39-54.

Kock, N. (2005). Media richness or media naturalness? The evolution of our biological communication apparatus and its influence on our behavior toward e-communication tools. IEEE transactions on professional communication, 48(2), 117-130.

Kock, N. (2009). Information systems theorizing based on evolutionary psychology: an interdisciplinary review and theory integration framework. Mis Quarterly, 395-418.

König, C. J., & Langer, M. (2022). Machine learning in personnel selection. In Handbook of Research on Artificial Intelligence in Human Resource Management (pp. 149-167). Edward Elgar Publishing.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences of the United States of America. 110(15), 5802-5805.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174.

Langer, M., König, C. J., & Krause, K. (2017). Examining digital interviews for personnel selection: Applicant reactions and interviewer ratings. International Journal of Selection and Assessment, 25(4), 371-382.

Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. International Journal of Selection and Assessment, 27(3), 217-234.

Latham, G. P. (1989). The reliability, validity, and practicality of the situational interview. In R. W. Eder & G. R. Ferris (Eds.), The employment interview: Theory, research, and practice (pp. 169–182). Sage Publications, Inc.

Latu, I. M., Mast, M. S., & Stewart, T. L. (2015). Gender biases in (inter) action: The role of interviewers' and applicants' implicit and explicit stereotypes in predicting women's job interview outcomes. Psychology of Women Quarterly, 39(4), 539-552.

Laurano, M. (2015). The True Cost of a Bad Hire. Brandon Hall Group Research Brief. https://b2b-assets.glassdoor.com/the-true-cost-of-a-bad-hire.pdf. Retrieved 28.06.2022.

Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. Journal of research in personality, 42(4), 914-932.

Letzring, T. D., Colman, D. E., Krzyzaniak, S. L., & Roberts, B. W. (2020). Realistic Accuracy Model. In B. J. Carducci & C. N. Nave (Eds.), The Wiley Encyclopaedia of Personality and Individual Differences, Volume 1, Models and Theories (pp. 341-349). John Wiley & Sons Ltd, New Jersey, USA.

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. Journal of personality and social psychology, 91(1), 111.

Leutner, F., Akhtar, R., & Chamorro-Premuzic, T. (2002). The Future of Recruitment. Using the New Science of Talent Analytics to Get Your Hiring Right. Emerald Publishing, United Kingdom.

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. Personnel Psychology, 67(1), 241-293.

Levine, S. P., & Feldman, R. S. (2002). Women and men's nonverbal behavior and self-monitoring in a job interview setting. Applied HRM Research, 7(1), 1-14.

Liden, R. C., Martin, C. L., & Parsons, C. K. (1993). Interviewer and applicant behaviors in employment interviews. Academy of Management Journal, 36, 372-386.

Linzer, M., Griffiths, E. P., & Feldman, M. D. (2022). Responding to the Great Resignation: Detoxify and Rebuild the Culture. Journal of General Internal Medicine, 1-2.

Lochner, K., & Preuss, A. (2018). Digitales Recruiting. Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO), 49(3), 193-202.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90-103.

Lorenzo, G. L., Biesanz, J. C., & Human, L. J. (2010). What is beautiful is good and more accurately understood: Physical attractiveness and accuracy in first impressions of personality. Psychological science, 21(12), 1777-1782.

Lüders, A. (2020). Nutzbarkeit visueller Hinweisreize in asynchronen Videointerviews: Eine Folgeuntersuchung (Bachelor's Thesis). Fachhochschule Westküste, Heide, Germany.

Luft, J., & Ingham, H. (1955). The Johari window: A graphic model of interpersonal awareness. Paper presented at the Proceedings of the Western Training Laboratory in Group Development. UCLA Extension Office, Los Angeles, USA.

Lukacik, E. R., Bourdage, J. S., & Roulin, N. (2022). Into the void: A conceptual model and research agenda for the design and use of asynchronous video interviews. Human Resource Management Review, 32(1), 100789.

Macan, T. (2009). The employment interview: A review of current studies and directions for future research. Human Resource Management Review, 19(3), 203-218.

McAbee, S. T., & Connelly, B. S. (2016). A Multi-Rater Framework for Studying Personality: The Trait-Reputation-Identity Model. Psychological Review, 123(5), 569-591.

McColl, R., & Michelotti, M. (2019). Sorry, could you repeat the question? Exploring video-interview recruitment practice in HRM. Human Resource Management Journal, 29(4), 637-656.

McConnell, B. (2021). Reading candidates body language in a virtual job interview. recruiteeblog. https://recruitee.com/articles/reading-candidate-body-language-in-a-virtual-job-interview. Retrieved 09.09.2022.

McCrae, R. R., & Costa, P. T., Jr. (2008). The five-factor theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), Handbook of personality: Theory and research (pp. 159–181). The Guilford Press.

McCrae, R. R., & John, O. P. (1992). An introduction to the Five-Factor Model and its applications. Journal of Personality, 60(2), 175-205.

Medina-Garrido, J. A., Biedma-Ferrer, J. M., & Ramos-Rodríguez, A. R. (2017). Relationship between work-family balance, employee well-being and job performance. Academia Revista Latinoamericana de Administración.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. Journal of Personality and Social Psychology, 90(5), 862–877.

Mejia, C., & Torres, E. N. (2018). Implementation and normalization process of asynchronous video interviewing practices in the hospitality industry. International Journal of Contemporary Hospitality Management, 30(2), 685-701.

Melchers, K. G., Bösser, D., Hartstein, T., & Kleinmann, M. (2012). Assessment of situational demands in a selection interview: Reflective style or sensitivity?. International Journal of Selection and Assessment, 20(4), 475-485.

Meltzer, R. (2020). How to conduct video interviews: 7 tips for employers. TechTarget Tip. https://www.techtarget.com/searchhrsoftware/tip/How-to-conduct-video-interviews-7-tips-for-employers. Retrieved 10.09.2022.

Microsoft Corporation (2021). The Next Great Disruption is Hybrid Work – Are We Ready? 2021 Work Trend Index: Annual Report.

Mirowska, A., & Mesnet, L. (2022). Preferring the devil you know: Potential applicant reactions to artificial intelligence evaluation of interviews. Human Resource Management Journal, 32(2), 364-383.

Motowidlo, S. J. (2003). Job performance. Handbook of psychology: Industrial and organizational psychology, 12(4), 39-53.

Murphy, N. A., Hall, J. A., Ruben, M. A., Frauendorfer, D., Schmid Mast, M., Johnson, K. E., & Nguyen, L. (2019). Predictive validity of thin-slice nonverbal behavior from social interactions. Personality and Social Psychology Bulletin, 45(7), 983-993.

Mwangi, B., Tian, T. S., & Soares, J. C. (2014). A review of feature reduction techniques in neuroimaging. Neuroinformatics, 12(2), 229–244.

National Careers Service (n.d.). Video interviews: how to do well. National Careers Service Careers Advice. https://nationalcareers.service.gov.uk/careers-advice/how-to-do-well-in-video-interviews. Retrieved 10.09.2022.

Nederström, M., & Salmela-Aro, K. (2014). Self-other agreement of personality judgments in job interviews: Exploring the effects of trait, gender, age and social desirability. Scandinavian Journal of Psychology, 55(5), 520-526.

New York City Council (2021). Automated employment decision tools. Int 1894-2020, Law No. 2021/144. Committee on Technology. New York City, USA.

Nguyen, L. S. (2015). Computational analysis of behavior in employment interviews and video resumes (Doctoral dissertation). Swiss Federal Institute of Technology Lausanne, Lausanne, Switzerland.

Nguyen, L. S., & Gatica-Perez, D. (2015). I would hire you in a minute: Thin slices of nonverbal behavior in job interviews. In Proceedings of the 2015 ACM on international conference on multimodal interaction (pp. 51-58).

Nielsen, J., Clemmensen, T., & Yssing, C. (2002). Getting access to what goes on in people's heads?: reflections on the think-aloud technique. In Proceedings of the second Nordic conference on Human-computer interaction (pp. 101-110). ACM.

Obermann, C. (2018). Assessment Center: Entwicklung, Durchführung, Trends. Mit neuen originalen AC-Übungen (6th Ed.). Springer, Köln, Germany.

Oliffe, J. L., Kelly, M. T., Gonzalez Montaner, G., & Yu Ko, W. F. (2021). Zoom interviews: benefits and concessions. International Journal of Qualitative Methods, 20, 16094069211053522.

Paunonen, S. V. (1989). Consensus in personality judgments: Moderating effects of target-rater acquaintanceship and behavior observability. Journal of personality and social psychology, 56(5), 823.

Peeters, H., & Lievens, F. (2006). Verbal and nonverbal impression management tactics in behavior description and situational interviews. International journal of selection and assessment, 14(3), 206-222.

Pingitore, R., Dugoni, B. L., Tindale, R. S., & Spring, B. (1994). Bias against overweight job applicants in a simulated employment interview. Journal of applied psychology, 79(6), 909.

Poh, W. Y. F. (2015). Evaluating candidate performance and reaction in one-way video interviews (Doctoral dissertation). San Francisco State University, San Francisco, USA.

Powell, D. M., & Bourdage, J. S. (2016). The detection of personality traits in employment interviews: Can “good judges” be trained?. Personality and Individual Differences, 94, 194-199.

Qiu, L., Lu, J., Yang, S., Qu, W., & Zhu, T. (2015). What does your selfie say about you? Computers in Human Behavior, 52, 443–449.

Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. Proceedings of the National Academy of Sciences, 114(41), 10870-10875.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. Academy of Management Review, 46(1), 192-210.

Rasipuram, S., & Jayagopi, D. B. (2016). Asynchronous video interviews vs. face-to-face interviews for communication skill measurement: a systematic study. In Proceedings of the 18th ACM International Conference on Multimodal Interaction. ACM.

Rentfrow, P. J., & Gosling, S. D. (2006). Message in a ballad: The role of music preferences in interpersonal perception. Psychological Science, 17(3), 236–242.

Robert Half Talent Solutions (2021). Video Interview Tips: A Job Candidate’s Checklist. The Robert Half Blog. https://www.roberthalf.com/blog/job-interview-tips/screen-time-how-to-nail-your-next-video-interview. Retrieved 09.09.2022.

Robert Walters Group (n.d.). Video interviews spike by 67% - according to recruitment firm. Robert Walters News. https://www.robertwalters.co.uk/news/video-interviews-spike.html. Retrieved 27.09.2022.

Roche, J. M., & Arnold, H. S. (2018). The effects of emotion suppression during language planning and production. Journal of Speech, Language, and Hearing Research, 61(8), 2076–2083.

Rogers, K. H., & Biesanz, J. C. (2015). Knowing versus liking: Separating normative knowledge from social desirability in first impressions of personality. Journal of Personality and Social Psychology, 109, 1105–1116.

Roth, P. L., & Huffcutt, A. I. (2013). A meta-analysis of interviews and cognitive ability: Back to the future?. Journal of Personnel Psychology, 12(4), 157.

Rotundo, M. (2002). Defining and measuring individual level job performance: A review and integration. Journal of Applied Psychology, 90(5), 225-254.

Roulin, N. (2022). The psychology of job interviews. Second Edition. Routledge.

Roulin, N., Bangerter, A., & Levashina, J. (2015). Honest and deceptive impression management in the employment interview: Can it be detected and how does it impact evaluations? Personnel Psychology, 68(2), 395-444.

Rudolph, C. W., Wells, C. L., Weller, M. D., & Baltes, B. B. (2009). A meta-analysis of empirical studies of weight-based bias in the workplace. Journal of Vocational Behavior, 74(1), 1-10.

Saarijärvi, M., & Bratt, E. L. (2021). When face-to-face interviews are not possible: tips and tricks for video, telephone, online chat, and email interviews in qualitative research.

Salgado, J. F., & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. European Journal of Work and Organizational Psychology, 11(3), 299-324.

Schlegel, K., Boone, R. T., & Hall, J. A. (2017). Individual differences in interpersonal accuracy: A multi-level meta-analysis to assess whether judging other people is one skill or many. Journal of Nonverbal Behavior, 41, 103–137.

Schmid Mast, M., Bangerter, A., Bulliard, C., & Aerni, G. (2011). How accurate are recruiters' first impressions of applicants in employment interviews?. International journal of selection and assessment, 19(2), 198-208.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological bulletin, 124(2), 262.

Schmidt, F. L., Oh, I. S., & Shaffer, J. A. (2016). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research findings. Working Paper available on Research Gate.

Schneider, L., Powell, D. M., & Roulin, N. (2015). Cues to deception in the employment interview. International Journal of Selection and Assessment, 23(2), 182-190.

Schwab, K. & Zahidi, S. (2020). The Future of Jobs Report 2020. World Economic Forum.

Segalin, C., Celli, F., Polonio, L., Kosinski, M., Stillwell, D., Sebe, N., Cristani, M., & Lepri, B. (2017). What your Facebook profile picture reveals about your personality. In Proceedings of the 25th ACM international conference on Multimedia (pp. 460–468). Association for Computing Machinery.

Sherman, R. A., Nave, C. S., & Funder, D. C. (2012). Properties of persons and situations related to overall and distinctive personality-behavior congruence. Journal of Research in Personality, 46(1), 87-101.

Shukla, V., Pandiya, B., Gupta, S., & Prashar, S. (2022). The Great Resignation: An Empirical Study on Employee Mass Resignation and its Associated Factors.

Sonnentag, S., Volmer, J., & Spychala, A. (2008). Job performance. The Sage handbook of organizational behavior, 1, 427-447.

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. The American Journal of Psychology, 15(2), 201-292.

Speer, A. B. (2018). Quantifying with words: An investigation of the validity of narrative-derived performance scores. Personnel Psychology, 71(3), 299-333.

Spence, M. 1973. Job market signaling. Quarterly Journal of Economics, 87: 355-374.

Spirito Dalgin, R., & Bellini, J. (2008). Invisible disability disclosure in an employment interview: Impact on employers' hiring decisions and views of employability. Rehabilitation Counseling Bulletin, 52(1), 6-15.

Springbett B (1958) Factors affecting the final decision in the employment interview. Can J Psychol 12(1):13.

Stamarski, C. S., & Son Hing, L. S. (2015). Gender inequalities in the workplace: the effects of organizational structures, processes, practices, and decision makers' sexism. Frontiers in psychology, 6, 1400.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT Approach to Constructing and Scoring Pairwise Preference Items Involving Stimuli on Different Dimensions: The Multi-Unidimensional Pairwise-Preference Model. Applied Psychological Measurement, 29(3), 184-203.

Straus, S. G., Miles, J. A., & Levesque, L. L. (2001). The effects of videoconference, telephone, and face-to-face media on interviewer and applicant judgments in employment interviews. Journal of management, 27(3), 363-381.

Sun, J., & Vazire, S. (2019). Do people know what they're like in the moment?. Psychological science, 30(3), 405-414.

Swider, B. W., Barrick, M. R., & Harris, T. B. (2016). Initial impressions: What they are, what they are not, and how they influence structured interview outcomes. Journal of Applied Psychology, 101, 625–638. https://doi.org/10.1037/apl0000077.

TestGorilla (n. d.). An easy-to-use guide to video interviewing for HR teams. https://www.testgorilla.com/blog/video-interviewing-guide-hr/. Retrieved 09.09.2022.

Thorndike, E. L. (1920). A constant error in psychological ratings. Journal of applied psychology, 4(1), 25.

Toldi, N. L. (2011). Job applicants favor video interviewing in the candidate-selection process. Employment Relations Today, 38(3), 19-27.

Torres, E. N., & Gregory, A. (2018). Hiring manager's evaluations of asynchronous video interviews: The role of candidate competencies, aesthetics, and resume placement. International Journal of Hospitality Management, 75, 86–93.

Torres, E. N., & Mejia, C. (2017). Asynchronous video interviews in the hospitality industry: Considerations for virtual employee selection. International Journal of Hospitality Management, 61, 4-13.

Tourangeau, R., Steiger, D. M., & Wilson, D. (2002). Self-administered questions by telephone – Evaluating interactive voice response. Public Opinion Quarterly, 66(2), 265-278.

University of Massachusetts Global Administration (n.d.). 7 Video interview tips that can help you stand out to employers. https://www.umassglobal.edu/news-and-events/blog/video-interview-tips. Retrieved 09.09.2022.

van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. MIS Quarterly, 45(3).

Van Den Haak, M., De Jong, M., & Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. Behavior & information technology, 22(5), 339-351.

van Esch, P., Black, J. S., & Arli, D. (2021). Job candidates' reactions to AI-enabled job application processes. AI and Ethics, 1(2), 119-130.

Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly, R. R. (2006). Forced-choice personality tests: A measure of personality and cognitive ability. Human Performance, 19, 175-189.

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. Journal of personality and social psychology, 98(2), 281-300.

Vazire, S., & Carlson, E. N. (2011). Others sometimes know us better than we know ourselves. Current Directions in Psychological Science, 20(2), 104-108.

VidCruiter (n.d.). Where Did Video Hiring Come From and Where is it Going?. https://vidcruiter.com/video-interviewing/history-of-video-interview/. Retrieved 27.09.2022.

Vijay, R. S., Shubham, K., Renier, L. A., Kleinlogel, E. P., Mast, M. S., & Jayagopi, D. B. (2021). An Opportunity to Investigate the Role of Specific Nonverbal Cues and First Impression in Interviews using Deepfake Based Controlled Video Generation. In Companion Publication of the 2021 International Conference on Multimodal Interaction (pp. 148-152).

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. International Journal of Selection and Assessment, 8(4), 216-226.

Waldman, J. D., Kelly, F., Arora, S., & Smith, H. L. (2010). The shocking cost of turnover in health care. Health care management review, 35(3), 206-211.

Wall, H. J., & Campbell, C. (2021). Accuracy of Personality Trait Judgments based on Environmental and Social Media Cues. In T.D. Letzring, & J. S. Spain (Eds.), The Oxford handbook of accurate personality judgment (pp. 219-234). Oxford University Press, Oxford, United Kingdom.

Wall, H. J., Taylor, P. J., Dixon, J., Conchie, S. M., & Ellis, D. A. (2013). Rich contexts do not always enrich the accuracy of personality judgments. Journal of Experimental Social Psychology, 49(6), 1190-1195.

Weltman, B. (2022). How Much Does an Employee Cost You?. U.S. Small Business Administration. https://www.sba.gov/blog/how-much-does-employee-cost-you. Retrieved 01.08.2022.

Wilhelmy, A., Roulin, N., & Wingate, T. G. (2021). Does it take two to tango? Examining how applicants and interviewers adapt their impression management to each other. Journal of Business and Psychology, 36(6), 1053-1076.

Wolgast, S., Björklund, F., & Bäckström, M. (2018). Applicant ethnicity affects which questions are asked in a job interview: The role of expected fit. Journal of Personnel Psychology, 17(2), 66.

Appendix

I – extract from the ADEPT User Guide Documentation, slide 5 – 10

ADEPT-15 Model

Style	Aspects	Brief Definition	Definition	Big Five Dimension
Teamwork Style	Cooperativeness	Cooperative, courteous, and trusting	This personality aspect reflects the extent to which someone is cooperative and trusting. People who score high tend to be team oriented and accommodating, but can sometimes be taken advantage of by others. Those who score low tend to be less interested in teamwork, but are also more independent-minded.	Agreeableness
	Sensitivity	Compassionate, caring, and understanding	This personality aspect reflects the extent to which someone is compassionate, caring, and understanding. Those who score high tend to be warmhearted, patient, and tolerant, but may have difficulty providing negative feedback or being firm with others. Those who score low tend to be stoic and tough-minded, but also candid and direct.	Agreeableness
	Humility	Modest, genuine, and selfless	This aspect of personality measures the extent to which someone is modest and genuine. High scorers tend to be humble and unselfish; but they may be less effective in advocating for their own interests. Low scorers are proud, cunning, shrewd, and can be manipulative; but are also bold and can be proficient at managing situations requiring tact and posturing.	N/A

ADEPT-15 Model

Style	Aspects	Brief Definition	Definition	Big Five Dimension
Task Style	Drive	Productive, proactive, and dependable	This personality aspect reflects the extent to which someone is proactive, and persistent. Those who score high tend to be reliable, hard working, and accountable, but may get overly focused on narrow goals and can be seen as rigid. Those who score low tend to be reactive and less deadline-oriented, but can shift more easily from goal to goal.	Conscientiousness
	Structure	Planful, detail-oriented, and rule-conscious	This personality aspect reflects the extent to which someone is planful, detail-oriented, and rule-conscious. High scorers tend to be careful, safe, and orderly, but also perfectionists. Low scorers tend to be disorganized and easily bored, yet can find innovative ways to handle problems and are more likely to focus on the big picture.	Conscientiousness

ADEPT-15 Model

Style	Aspects	Brief Definition	Definition	Big Five Dimension
Emotional Style	Composure	Composed, calm, and relaxed	This personality aspect reflects the extent to which someone is composed, calm, and relaxed. High scorers tend to be tranquil, retrained, and calm under pressure, but can seem aloof and detached. Low scorers tend to be impulsive and excitable; but also demonstrate passion, excitement, and enthusiasm.	Emotional Stability
	Positivity	Happy, optimistic, and resilient	This personality aspect reflects the extent to which someone is happy, optimistic, and resilient. High scorers tend to be hopeful and positive, but may downplay or disregard potential problems. Low scorers can be pessimistic and overwhelmed with obstacles, but tend to be more pragmatic. Low scorers also are effective advocates for unpopular decisions.	Emotional Stability
	Awareness	Reflective and self-aware	This aspect of personality measures the extent to which someone is reflective and self-aware. High scorers are introspective and know their own strengths and weaknesses, but may be self-absorbed. Low scorers have a static self concept and are resistant to feedback, yet are less concerned with or care what others think about them.	N/A

7 AON Hewitt

ADEPT-15 Model

Style	Aspects	Brief Definition	Definition	Big Five Dimension
Interaction Style	Assertiveness	Assertive, decisive, and competitive	This aspect reflects the extent to which someone is assertive, decisive, and competitive. High scorers are persuasive and bold, but can be confrontational and aggressive. Low scorers are less concerned with winning and are more cautious when making decisions. Also, they prefer to avoid conflict and may give into others too easily.	Extraversion
	Liveliness	Outgoing, energetic, and confident	This aspect of personality focuses on the extent to which someone is outgoing, energetic, and socially confident. High scores tend to be sociable and friendly, though they may be rambunctious and attention seeking. Low scorers tend to be more reserved and quiet, but also more private and unlikely to offend to others.	Extraversion

8 AON Hewitt

ADEPT-15 Model

Style	Aspects	Brief Definition	Definition	Big Five Dimension
Adaptation Style	Conceptual	Thoughtful and curious	This aspect of personality measures the extent to which someone is conceptual and intellectually curious. High scorers tend to be inquisitive and philosophical, but may be overly abstract and unrealistic. Low scorers tend to be conventional with less curiosity, as well as more concrete and practical.	Openness
	Flexibility	Flexible, adaptable, and open-minded	This aspect of personality measures the extent to which someone is flexible, adaptable, and open-minded. High scorers tend to be open to new ideas and experiences, but may come off as inconsistent or indecisive. Low scorers may be inflexible and set in their ways, but more predictable as they seek tried-and-true approaches.	Openness
	Mastery	Learning-oriented and improvement-focused	This aspect of personality measures the extent to which someone is learning-oriented and improvement-focused. High scorers tend to be focused on self-development, practice, and the belief that others can improve; though may be unrealistic in their views of others or their own potential. Low scorers are less concerned with continual self-development, and believe people do not often change much, but they can focus on getting done what is needed	N/A

9 AON Hewitt

ADEPT-15 Model

Style	Aspects	Brief Definition	Definition	Big Five Dimension
Achievement Style	Ambition	Ambitious and goal-directed	This aspect of personality measures the extent to which someone is ambitious and goal-directed. High scorers are relentless in their pursuits, but can be obsessive and are rarely satisfied. They may also pursue individual goals in lieu of team goals. Low scorers are satisfied with their current status and often have a good work-life balance.	N/A
	Power	Controlling, directive, and motivated to lead	This aspect of personality measures the extent to which someone is controlling, directive, and motivated to lead. High scorers tend to be interested in leadership, control, and influence. Low scorers tend to be team players, lead by example, and willing to let others to take control.	N/A

10 AON Hewitt

II – extract from the vidAssess User Manual, slides 4 – 20

vidAssess: Core Concepts Overview

- Asynchronous Video Interview: Creates individual interviews for specific target groups/job positions
- Fully customisable: define competency model, create custom question pools, embed images and videos, define preparation time, recording time, number of retries.
- Interview can contain up to 10 questions
- Completed via desktop or mobile browser function (Chrome, Firefox, Edge or Opera)
- Recruiters: Single or multiple recruiters can be assigned to rate videos
- Results: Reporting using different levels (e.g. Single Average Score, Area-Score, Criteria-Score, Question-Score)
- Rating: Based on defined competency model – standardised

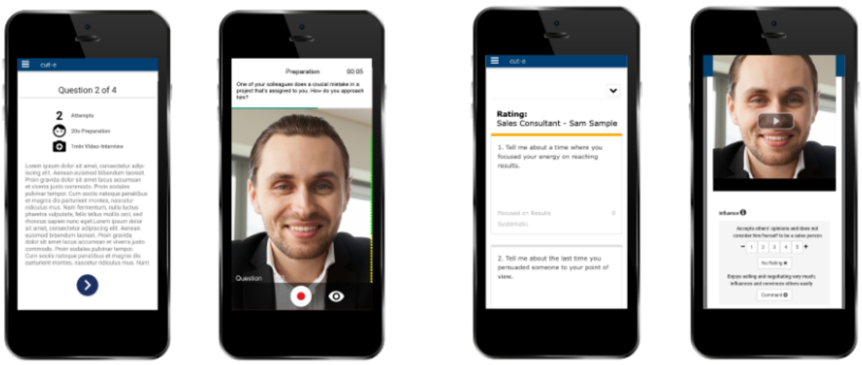
smartPredict: vidAssess can be presented as a second task besides a smartPredict task battery

chatAssess: vidAssess cannot be used within chatAssess (warp)

ATS-Integration: integrated as single instrument or within a candidate Hub Integration. vidAssess API (Application Programming Interface) for rating integrations

AON Empower Results® 4

vidAssess: Video Interviewing



Candidate view

Rater view

AON Empower Results®

Introduction to vidAssess Setup

Getting Started

Before vidAssess can be used, a couple of activation steps have to be done:

1. The instrument 470_vidAssess must be activated within partner and client account.
2. The vidAssess accounting level must be activated on partner and client level.
3. Furthermore, necessary vidAssess permissions must be assigned to the respective admin.
4. Specific candidate languages must be activated manually after the translation is completed.

Activation

- 470_vidAssess instrument and the accounting level needs to be activated - Completed by Ops Manager on request.
- Partner admins must activate the instrument and the accounting level for each client account – under 'external instruments'
- Activate the vidAssess accounting level using the lock. A green check mark indicates that the accounting level is activated.

Permissions

- Multiple permissions for specific purposes are available. Please see appendices for permissions definitions
- Click on 'Actions' and then 'Administrator' on the Home screen or inside a partner/client/project. On the 'Administrators page', click on 'Permissions' to assign permissions to the respective Administrator.

Translations

Regarding language availability, translations, or missing texts, please contact translations.assessment@aon.com There are 3 different areas of vidAssess specific texts:

- System Texts: Systems – mapTQ > Text Content > Text Content
  - TextIDs from 470.000 to 480.000
- Player Texts: Instruments > 10_vidAssess > vidAssessPlayer
- Report Texts: reports > Text content > vidAssess – Competency Report > text_content

Accounting

For the usage of vidAssess, the following units will be charged:

- 15 units (vidAssess level) when a candidate starts the first question and goes past the Introduction Page
- Additional 20 units per candidate (screening level) when they complete the interview
- The reports and data exports do not charge additional units

AON Empower Results®

6

vidAssess Setup

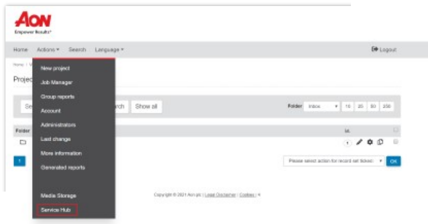
In order to access the vidAssess Setup within the vidAssess Menu, the vidAssess Setup permission is required. Activate the instrument 470 vidAssess and the vidAssess accounting level on the Partner/Client.

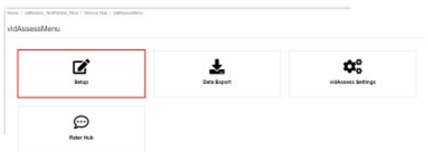
Setup

1. Set up a Competency Model that you want to use for a specific job or that is used in the company in general.
2. Set up a Question Pool that contains all possible interview questions
3. Thirdly, set up an Interview for a specific Job or a specific situation. You will be able to choose up to 10 questions from the Question Pool that you linked with the Interview. Also, you will be able to customize welcome, introduction and closing texts, to adapt the interview appropriately to the expected candidates or the situation.

To start the setup process, click 'Actions', then 'Service Hub' and afterwards on 'VidAssess' on Partner or Client level. Finally, click 'Setup'.

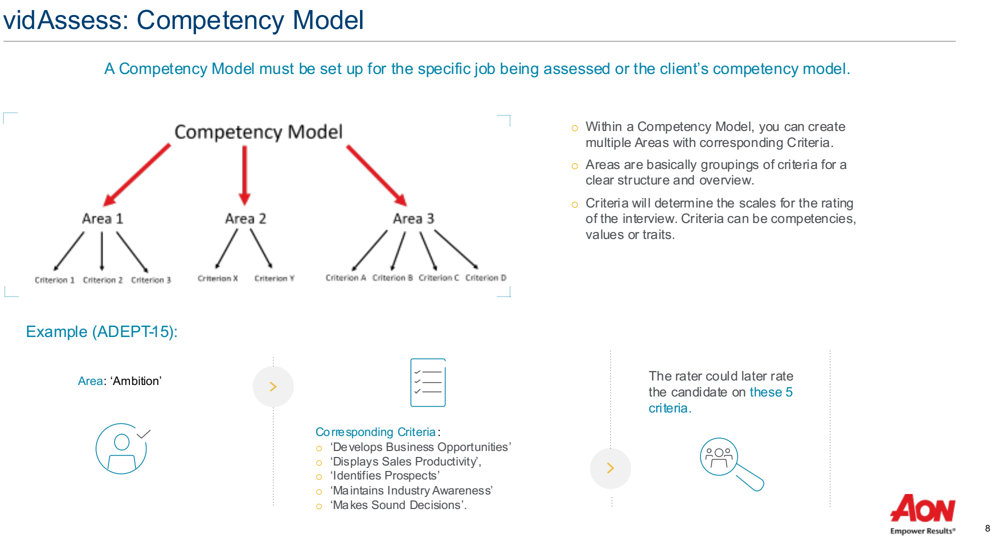
Note, elements set up on Partner level can later be enabled for respective clients but not vice versa.

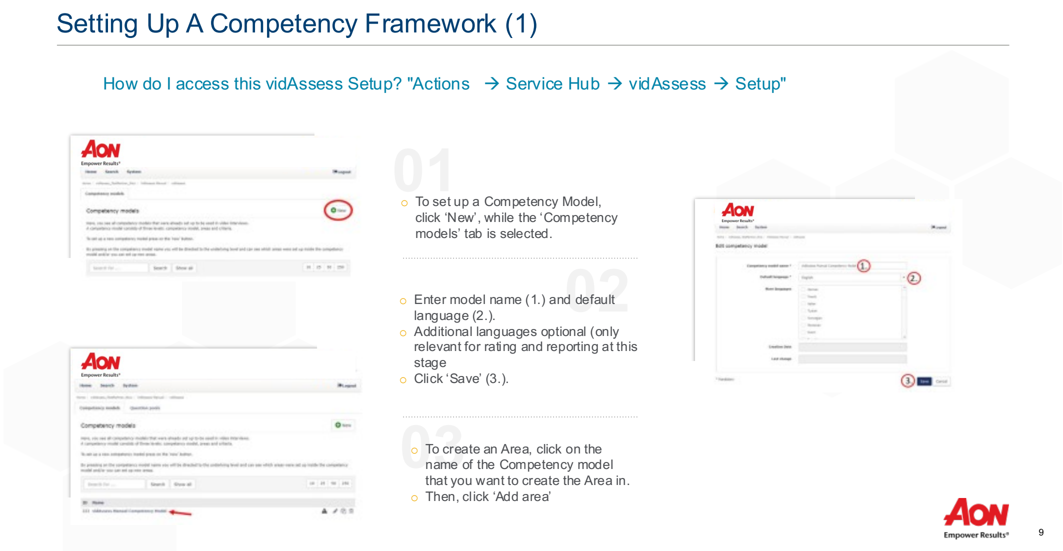




AON Empower Results®

7


vidAssess: Competency Model
A Competency Model must be set up for the specific job being assessed or the client's competency model.
Competency Model
Area 1
Area 2
Area 3
Criterion 1
Criterion 2
Criterion 3
Criterion X
Criterion Y
Criterion A
Criterion B
Criterion C
Criterion D
Within a Competency Model, you can create multiple Areas with corresponding Criteria.
Areas are basically groupings of criteria for a clear structure and overview.
Criteria will determine the scales for the rating of the interview. Criteria can be competencies, values or traits.
Example (ADEPT-15):
Area: 'Ambition'
Corresponding Criteria:
'Develops Business Opportunities'
'Displays Sales Productivity',
'Identifies Prospects'
'Maintains Industry Awareness'
'Makes Sound Decisions'.
The rater could later rate the candidate on these 5 criteria.
AON
Empower Results
8


Setting Up A Competency Framework (1)
How do I access this vidAssess Setup? "Actions → Service Hub → vidAssess → Setup"
01
To set up a Competency Model, click 'New', while the 'Competency models' tab is selected.
02
Enter model name (1.) and default language (2.).
Additional languages optional (only relevant for rating and reporting at this stage
Click 'Save' (3.).
03
To create an Area, click on the name of the Competency model that you want to create the Area in.
Then, click 'Add area'
AON
Empower Results
9

Setting Up A Competency Framework (2)

04

- Fill in the name of the Area in all previously activated languages. Then click 'Save'
- If areas are not needed, you can just create a 'dummy area' called 'criteria' or 'competencies'

05

- When the Area is set up, click on the name of it and then 'Add Criterion'

06

- You will need to fill in name and definition of the criterion as well as the minimum, mid and maximum description of the scale in all activated languages.

Aon Empower Results®

10

Question Pool: Question Types

Standard Question is the question type used to present text, images and videos to the candidate. They have the option to be linked with criteria and a comment box for the rating.

Audio-only Questions will only record candidates' microphone and no video. Apart from this aspect, this question type is identical to standard questions.

Virtual Case Study can be used to set up multiple sections of content with text, graphics, images and videos (relevant questions need to be set up separately).

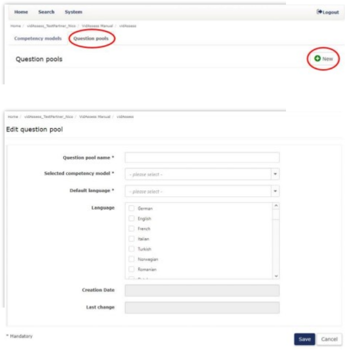
Video question can be used to display questions in video format that is, unlike in Standard Questions, detached from the recording page (can also be linked with criteria and comment boxes for the rating process)

Overall scales become part of the rating process. They serve as additional rating scales, not connected to any question and not visible to candidates

Aon Empower Results®

11

Setting Up a Question Pool



- Still in the setup (Actions → Service Hub → vidAssess → Setup), Question Pool creation becomes accessible as soon as a Competency Model is set up. The Question Pool serves as a collection of all questions that you might want to use at some point (for certain areas, certain positions, or for a certain project).
- You can create as many Question Pools as you like but note, that a specific Interview can only be linked to one Question Pool.

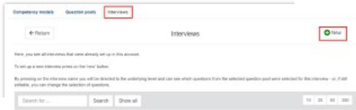
- Now, fill in the name, select a Competency Model (to be used for rating), and choose the languages in which you want to make the questions available in (for the candidates). When you're done, click 'Save'.
- Important: Once the Competency Model is in use in a Question Pool, it is no longer possible to edit the model
- Note, you can still copy a Competency Model and edit the copy.
- In order to use a Question Pool, it needs at least one Example Question and one Standard Question.
- Setting up an example question first is mandatory, and only Standard and Audio-only Questions can be created as the example should display the standard procedure of vidAssess. The setup of the example varies slightly from the subsequent questions (e.g., infinite attempts)

Question Pool: Setup

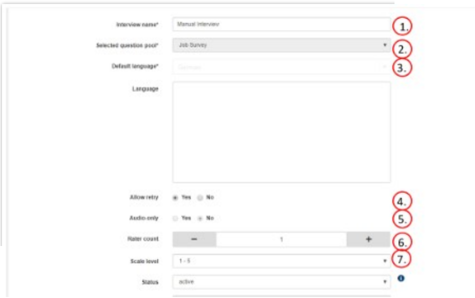
Question Type	Standard Question Audio-only Question	Virtual Case Study	Video Question	Overall Scale
Setup Guide	1. Set the Preparation Time Time before recording begins 2. Set the Response Time– minimum and maximum 3. Set number of retries available for each question 4. Enter a question name that serves as a label in the question pool 5. Enter the desired question text. Multiple HTML-functions are available Rating Settings: 1. Choose whether the question is supposed to be linked with certain criteria 2. If criteria are enabled, select any number of criteria from the linked Competency Model 3. Enable/disable a rating scale for the chosen criteria 4. Choose if you want to enable comment boxes	1. Select the number of sections that you want to fill with content (19 sections). 2. Specify the investigation time (30 seconds to 45 minutes). 3. Add the number of retries (0-20) 4. Name the Case Study and label all sections– Click 'Next' 5. Set up Instruction text 6. Fill in case study content using 'Case Study' tab **Note:** All questions regarding the Virtual Case Study's content need to be set up like any other question in vidAssess and separately from the Virtual Case Study.	1. Media– 2 ways of embedding videos ◦ YouTube videos hosted online & embedded via URL ◦ Videos uploaded from Media Storage & embedded in interview 2. Select investigation time (maximum time a candidate gets to read instructions– investigation time cannot be less than the length of the video 3. Select the amount of content retries 4. Add an instruction text 5. Add a YouTube video link/ Media Storage file 6. Following settings are identical to the standard question guide	Enables recruiters to rate candidates on more overall criteria. 1. Add a scale name that serves as a label within the question pool 2. Add a scale text, that serves as a scale description 3. Choose whether the scale is supposed to contain certain criteria 4. If criteria are enabled, select any number of criteria from the linked Competency Model that you want to associate with this specific question 5. Choose whether you want to enable a comment box for the rater 6. Decide whether a comment on this scale is mandatory or not Categorical Scales can also be set up.

Interview Setup

Once all prerequisites are created, interviews can now be setup and customised.



To start the interview setup, go to the "interviews" tab and click "New".



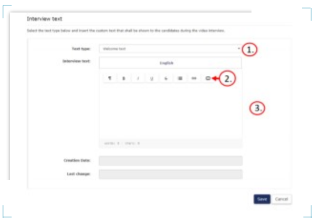
1. Name the Interview (visible to the candidate)
2. Select the underlying Question Pool
3. Select the Interview language(s) you want to be enabled for candidates
4. Allow or disable the retries defined earlier
5. Activate the audio-only setting in case you want to disable the recording of videos for the entire interview, only microphone will be recorded this way – deactivated per default, only change after careful consideration
6. Select how many raters are needed to complete a rating. Also, human ratings can be completely deactivated by setting the count to "no rating".
7. Select the scale on which the Interview will later be rated (from 1- 2 up to a 1-9-point-scale).

After choosing the relevant questions from the question pool, and all settings are chosen, you can set the interview status to "active". As soon as the interview is active, changes are no longer possible.

14

Customising Interview Texts

To add and customize the Interview Texts, click on the 'Interview text' tab. The purpose of Interview Texts is to customize the Interview and adapt displayed texts to the candidates or the situation. Each custom section is entirely optional. Click 'Add text' to set up new texts.



1. *Text Type* : Where the text will be shown (e.g., welcome text, closing text)
2. *Text Formatting:* Here you can adjust the format and embed media storage files, e.g., videos and images
3. *Interview Text :* Insert whatever you wish to display to candidates

All sections can include an embedded video/link

If no changes are made, texts will revert to default texts

Welcome Text

- First page of the interview
- An opportunity to provide a short introduction to the company and the role in question, or provide a general brand building video.

Introduction Text

- Appears after the example before the actual interview starts
- A final opportunity to offer candidates additional information before they launch the video interview recording

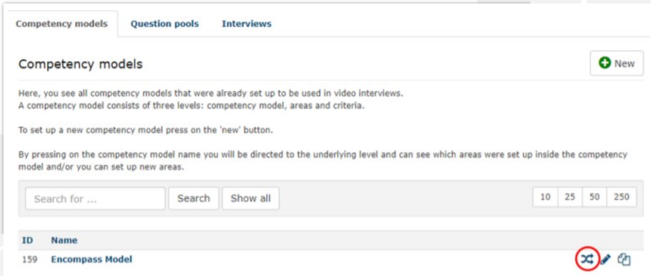
Closing Text

- Appears at the end of the interview
- This confirms completion of the assessment, and a thank you to the candidate for participating

15

Enabling Partner Elements on Client Level

In order to activate a vidAssess project for a client, you must enable partner elements on client level by using the Availability Tool



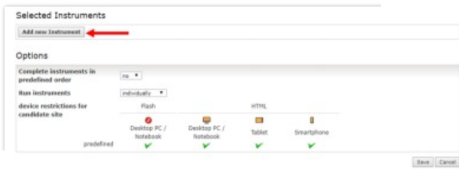
The arrow circled opens a drop-down menu with all clients.
Choose all clients that you want to activate the respective element on and click 'Save'

Aon Empower Results®

16

Adding an Interview to a Project

- To add vidAssess to a Project, go to the project tool and click on 'Instruments' next to the Project that you want vidAssess to be added to.
- Now, you will see a list of selected instruments. Above will be the 'Add new Instrument' button - click it



- vidAssess has its own tab, where all active Interviews are listed, which only appears if an active vidAssess Interview is available within that client.
- Select the Interview that you want to use in the Project. Note: You can only select one Interview per Project.
- Click 'Add', after selecting the Interview. When you return to the list of selected Instruments, click 'Save'.

Aon Empower Results®

17

vidAssess Settings

Access: 'Actions' → 'Service Hub' → 'vidAssess' → 'vidAssess Settings'.

01

vidAssess *Candidate View*

- Allows for sorting candidates by their vidAssess Task Process Status for a more convenient overview (simply by clicking 'vidAssess').
- In case of AI usage, an extra 'AI Status' column will appear

02

Quick Start Button

- Per default, the 'quickstart' button is set to 'always shown'
- By setting 'always hidden', the button will never be accessible for candidates and they will have to go through the example every time they (re-)start vidAssess
- By setting 'shown when started', the candidate will only have to go through the introduction and example when he starts his/her interview, but will be able to skip it in case of restarts

03

Storage

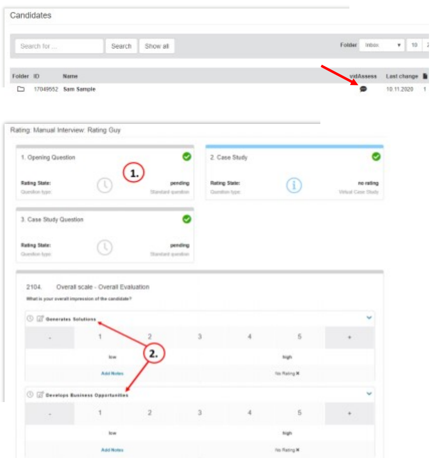
- This displays the storage location for candidate videos (only changeable with respective permission).

04

Send candidates with blocked or refused questions to rating

- Per default, candidates that got at least one blocked/refused question during their interview will still not be sent to the rating stage, when they complete their interview. Candidates will be marked on the candidate view with a '!' in their rating bubble. An Admin will have to either reset the respective blocked/refused question or manually complete the interview.
- If you want candidates with blocked/refused questions to automatically be sent to the rating stage, turn this setting to 'yes'

Rating Overview



The Rating Page is accessed in two ways:

- By clicking on a candidate's rater bubble on the candidate view, you will reach the Rating Overview.
- If the interview is not completed yet, the bubble will not show on the candidate view. In this case, view the task list and click the bubble there.

On the Rating Overview (a summary page of all raters and/or AI ratings), click 'assign to rating' and you will be listed as a rater. You will then be able to open the Rating Page.

- Each answer can be rated individually (**1**)
- If overall scales are set up, they will be found below (**2**)
- Once every scale is rated, press submit to finish rating

IV – Literature used to generate the initial visual cue list

Publication	**Visual cue category**
Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. Journal of personality and social psychology, 64(3), 431.	Face and Body
Batrinca, L. M., Mana, N., Lepri, B., Pianesi, F., & Sebe, N. (2011). Please, tell me about yourself: automatic personality assessment using short self-presentations. In Proceedings of the 13th international conference on multimodal interfaces (pp. 255-262). ACM.	Face and Body
Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of personality and social psychology*, *62*(4), 645.	Face and Body
Borkenau, P., & Liebler, A. (1995). Observable Attributesas Manifestations and Cues ofPersonality and Intelligence. Journal of Personality, 63(1), 1-25.	Face and Body
Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. Journal of personality and social psychology, 86(4), 599.	Face and Body
Burnett, J. R., & Motowidlo, S. J. (1998). Relations between different sources of information in the structured selection interview. Personnel Psychology, 51(4), 963-983.	Face and Body
Carney, D. R., Jost, J. T., Gosling, S. D., & Potter, J. (2008). The secret lives of liberals and conservatives: Personality profiles, interaction styles, and the things they leave behind. Political Psychology, 29(6), 807-840.	Face and Body
DeGroot, T., & Motowidlo, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. Journal of Applied Psychology, 84(6), 986.	Face and Body
Feiler, A. R., & Powell, D. M. (2015). Behavioral expression of job interview anxiety. Journal of Business and Psychology, 31(1), 155-171.	Face and Body
Grünberg, M., Mattern, J., Geukes, K., Küfner, A., Back, M., Brauner, E., Boos, M., & Kolbe, M. (2018). Assessing Group Interactions in	Face and Body

Personality Psychology: The Münster Behavior Coding-System (M-BeCoSy).	
Levine, S. P., & Feldman, R. S. (2002). Women and men's nonverbal behavior and self-monitoring in a job interview setting. Applied HRM Research, 7(1), 1-14.	Face and Body
Murphy, N. A. (2007). Appearing smart: The impression management of intelligence, person perception accuracy, and behavior in social interaction. Personality and Social Psychology Bulletin, 33(3), 325-339.	Face and Body
Murphy, N. A., Hall, J. A., Ruben, M. A., Frauendorfer, D., Schmid Mast, M., Johnson, K. E., & Nguyen, L. (2018). Predictive validity of thin-slice nonverbal behavior from social interactions. Personality and Social Psychology Bulletin.	Face and Body
Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality Judgments Based on Physical Appearance. Personality and Social Psychology Bulletin, 35(12), 1661–1671.	Face and Body
Nguyen, L. S. (2015). Computational analysis of behavior in employment interviews and video resumes (Doctoral dissertation, Ph. D. Dissertation. École Polytechnique Fédérale de Lausanne).	Face and Body
Parsons, C. K., & Liden, R. C. (1984). Interviewer perceptions of applicant qualifications: A multivariate field study of demographic characteristics and nonverbal cues. Journal of Applied Psychology, 69(4), 557.	Face and Body
Back, M. D., Schmukle, S. C., & Egloff, B. (2011). A closer look at first sight: Social relations lens model analysis of personality and interpersonal attraction at zero acquaintance. European Journal of Personality, 25(3), 225-238.	Appearance
Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of personality and social psychology*, *62*(4), 645.	Appearance
Borkenau, P., & Liebler, A. (1995). Observable Attributes as Manifestations and Cues of Personality and Intelligence.	Appearance
Journal of Personality, 63(1), 1-25.	
Nestler, S., Egloff, B., Küfner, A. C., & Back, M. D. (2012). An integrative lens model approach to bias and	Appearance

accuracy in human inferences: Hindsight effects and knowledge updating in personality judgments. Journal of personality and social psychology, 103(4), 689.	
Guntuku, S. C., Qiu, L., Roy, S., Lin, W., & Jakhetiya, V. (2015). Do others perceive you as you want them to?: Modeling personality based on selfies. In Proceedings of the 1st international workshop on affect & sentiment in multimedia (pp. 21-26). ACM.	Media Properties
Musil, B., Preglej, A., Ropert, T., Klasinc, L., & Babič, N. Č. (2017). What is seen is who you are: Are cues in selfie pictures related to personality characteristics?. Frontiers in psychology, 8, 82.	Media Properties
Qui, L., Lu, J., Yang, S., Qu, W., & Zhu, T. (2015). Was does your selfie say about you? Computers in Human Behavior, 52, 443-449.	Media Properties
Gosling, S. D., Craik, K. H., Martin, N. R., & Pryor, M. R. (2005). The personal living space cue inventory: An analysis and evaluation. Environment and Behavior, 37(5), 683-705.	Environment
Graham, L. T., & Gosling, S. D. (2011). Can the ambiance of a place be determined by the user profiles of the people who visit it?. In Fifth International AAAI Conference on Weblogs and Social Media.	Environment
Nguyen, L. S., Ruiz-Correa, S., Mast, M. S., & Gatica-Perez, D. (2017). Check out this place: Inferring ambiance from airbnb photos. *IEEE Transactions on Multimedia*, *20*(6), 1499-1511.	Environment
Qui, L., Lu, J., Yang, S., Qu, W., & Zhu, T. (2015). Was does your selfie say about you? Computers in Human Behavior, 52, 443-449.	Environment

V –Video Interview Questions in dataset 1

No	Question	Prompts	Workstyle	Aspects
1	Tell me about a time when you were planful and organized in achieving a goal or meeting a deadline.	What was the situation? Why was this important to you? How did you organize yourself? What was the result?	Task Style	Drive, Structure
2	Describe a time when you suggested a change in a process or procedure because you saw the potential for improvement.	What did you want to change and why was it necessary? How did you approach the change? What measures did you take? What was the result?	Adaption Style	Conceptual, Flexibility, Mastery
3	Describe a situation when you were particularly proud of your ability to lead others, even though you did not have formal authority.	Why and how did you take the lead? What abilities helped you to lead the others? What was the result of that situation?	Achievement Style	Ambition, Power
4	Tell me about a specific situation where you have taken proactive steps to build and maintain relationships with key colleagues or fellow students.	What steps did you take? How did the other person react? What did you do to maintain that relationship?	Interaction Style	Assertiveness, Liveliness
5	Describe a situation where you successfully established trust with someone who was previously resistant or skeptical	Why was the other person resistant? How did you react to the other person's concerns? How did you establish trust?	Teamwork Style	Cooperativeness, Sensitivity, Humility
6	Tell me about a time when you received constructive feedback and used it to improve your performance at work or school.	For what did you get feedback? What did you learn from it for the future? How did improve your performance?	Emotional Style	Composure, Positivity, Awareness

VI – Instruction for the pilot study to study 1

Instructions for Participants of Project Sherlock Thinking out Loud Pilot Study

In this study we are interested in what you see in video-interviews. You will be asked to verbally state visual cues as you see them in the interview you will be watching, following a thinking out loud approach.

A visual cue is any aspect of visual information present in the video interview. The videos are not manipulated and there are no specific items or behaviors to search for. Any visual cues you will see occur naturally, and thus anything you see can be mentioned by you.

Categories of visual cues and examples are listed below

Environment - e.g. “dark”, “messy shelf”, “art”

Face – e.g. “smile”, “looks at camera”, “nodding”

Body language – e.g. “hand movement”, “leans towards camera”

Appearance – e.g. “makeup”, “fancy clothes”, “dyed hair”

We ask you only focus on one category per video as stated in the instructions. You are free to mention as many cues as you notice per video including those stated as examples here should they appear in the video.

If you notice multiple cues at the same time, simply list them. You do not need to explain your mentions, only state the cue. We are only interested in whether cues are present or not.

Valid mention: “dyed hair”

Invalid mention: “I do not like this type of blond” or “the blond hair looks artificial”

Please make sure that all verbalizations are in English.

After a warm-up task allowing you to familiarize yourself with the process, you will watch an interview consisting of multiple separate videos. Mentions from the warm-up task will not be stored.

Multiple mentions across and within videos are allowed.

During this study your voice will be recorded using vidAssess. Each question in the vidAssess Interview used to record your voice corresponds to one video in the interview you will be watching. The category of visual cues to look out for will be stated in the instructions.

The video recording will not be used in this study, only the audio data. Please use the warm-up task to ensure your voice is recorded clearly.

VIII - Instruction for study 1

Instructions for Project Sherlock Thinking out Loud Study

In this study we are interested in your impressions of a candidate's personalities and how your form these impressions. You will be asked to watch vidAsses video interviews and rate candidates on ADEPT-15 workstyles.

The videos will be muted and your rating is to be based solely on visual information. We encourage you to use any visual cues you see in the categories of visual information listed below.

A visual cue is any aspect of visual information present in the video interview.

The videos are not manipulated and there are no specific items or behaviors to search for.

Any visual cues you will see occur naturally, and thus anything you see can be used for your judgement and be mentioned by you.

Categories of visual cues and examples are listed below

- Environment – examples: “dark”, “messy shelf”, “public location”
- Face – examples: “smile”, “looks at camera”, “nodding”
- Body language – examples: “hand movement”, “leans towards camera”
- Appearance – examples: “makeup”, “fancy clothes”, “dyed hair”
- Media Properties – examples: "face not fully visible", "person far away"

You will not need to explain your mentions, only state the cue. We are only interested in whether the presence or non-presence of certain visual cues has helped inform your personality rating.

- Valid mention: “dyed hair”
- Invalid mention: “I do not like this type of blond” or “the blond hair looks artificial”

Please make sure that all verbalizations are in English. During this study your voice will be recorded to allow for subsequent transcription of your mentions.

Prompts

Participants may be encouraged to recall and state more cues for a given personality dimension by using one of the following prompts. The prompts will be chosen depending on subjective fit to the study situation.

- Was there anything else you took into account for your rating?
- Did you notice anything else in the video?

ADEPT-15 Workstyles

Personality traits are best interpreted on a continuum. Low scores and high scores are neither inherently good nor bad. Both low and high scores have beneficial and undesirable implications for behavior. The scores reflect a likelihood of displaying certain behaviors.

Task Style

The task style is a broad measure of conscientiousness capturing one's approach to duties, responsibilities, and getting things done.

Consists of Drive and Structure.

Big5 equivalent: Conscientiousness

Teamwork Style

Teamwork Style is a broad assessment of agreeableness that describes how you approach relationships and how focused you are on the needs of others versus your own needs.

Consists of: Cooperativeness, Sensitivity, Humility.

Big5 equivalent: Agreeableness

Adaptation Style

Adaptation Style relates to a person's openness to experience and approach to learning and adapting to situations

Consists of: Conceptual, Flexibility, and Mastery

Big5 equivalent: Openness

Emotional Style

Your Emotional Style describes how you experience and react to feelings and you degree of self-awareness.

Consists of: Composure, Positivity, and Awareness

Big5 equivalent: Emotional Stability

Achievement Style

Need for achievement, including focus on career goals and influence over others is captured by Achievement Style.

Consists of Ambition, and Power

Big5 equivalent: N/A

Interaction Style

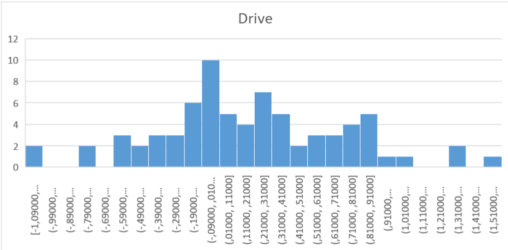
Interaction Style is a broad measure of extraversion that describes how much you seek out interaction with others and how you prefer to engage with them.

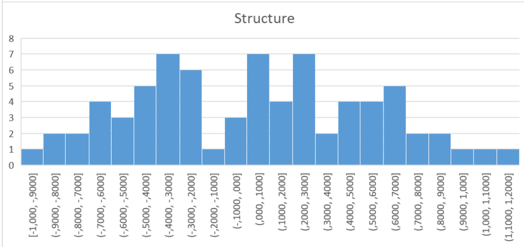
Consists of: Assertiveness and Liveliness

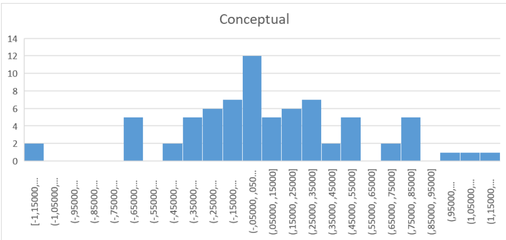
Big5 equivalent: Extraversion

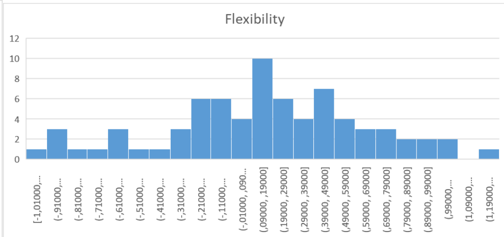
IX –

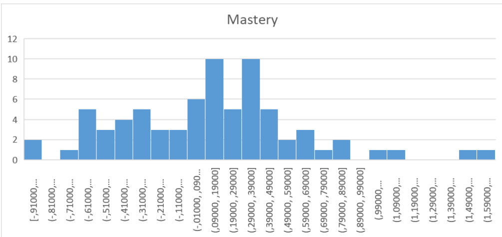
ADEPT-15 distribution per dimensions of dataset 1 subsample for study 1.5

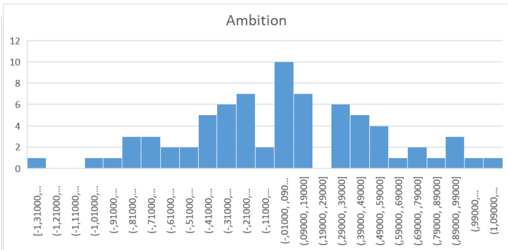


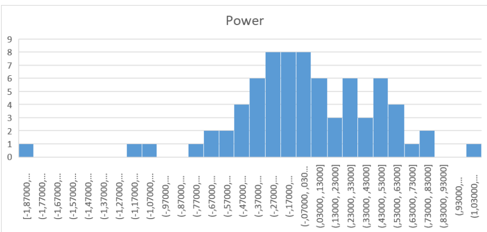


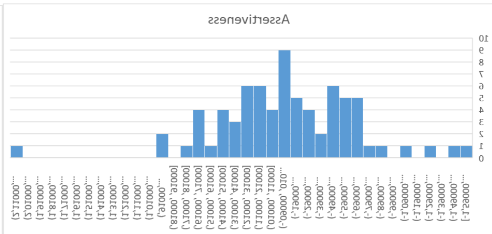


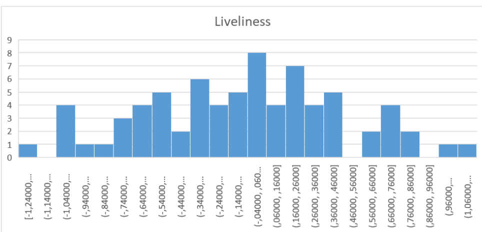


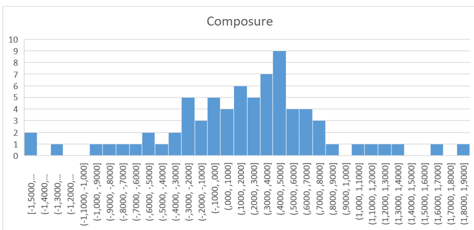


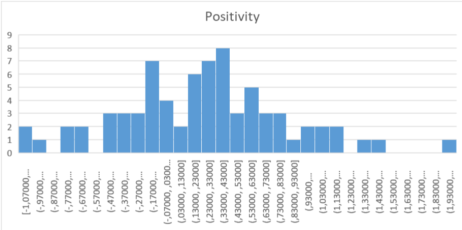


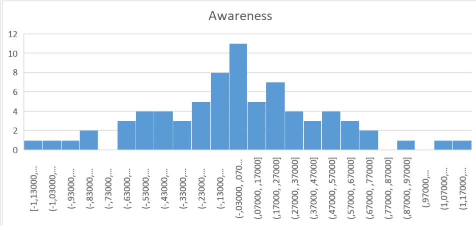


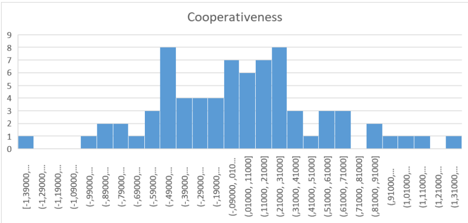


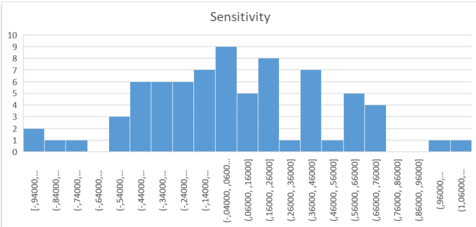


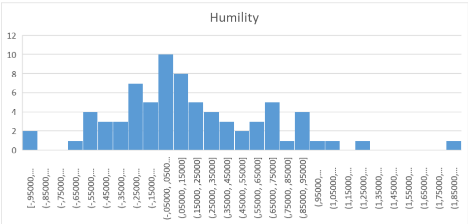



Positivity


Awareness


Cooperativeness


Sensitivity


Humility

X – Visual Cue Inventory

OCR-Version first, layouted version second

Target:____________

Judge:_____________

Date:_________

Time:_____________

Personality Ratings

	Low				High
Task Style	1	2	3	4	5
Adaptation Style	1	2	3	4	5
Achievement Style	1	2	3	4	5
Teamwork Style	1	2	3	4	5
Emotional Style	1	2	3	4	5
Interaction Style	1	2	3	4	5

Visual Cue Inventory A

FACE

Facial Expression

Friendly

Cheerful

Interested

Self-assured

Timid

Calm

Surprised

Skeptical

Arrogant-amused

Grumpy

Indifferent

Strong

Diverse

Rapidly changing

Forehead

Wrinkled

Other

Look

At camera

Away from camera sideways

Up

Down

Other

Eyes & Eyebrows

Rolling eyes

Heavy blinking

Wide open eyes

Closed eyes

Narrow eyes

Tinkles with eyelashes

Furrowed eyebrows

Raised eyebrows

Lips

Laughing

Smiling

Biting lips

Licking lips

Pressed lips

Plops lips

Duckface

Pout

Other

Mouth & Throat

Wide open mouth

Pauses

Swallowing hard

Fast mouth movements

Other

IMAGE QUALITY

Camera shaking

Other

BODY

Legs

Shaking legs

Legs crossed

Legs open

Other

Visual Cue Inventory B

BODY

Body Posture

Straight posture

Slouched posture

Relaxed posture

Maintained posture

Open posture

Stiff posture

Body tilted

Body turned away

Parallel body orientation

Leaning backward

Leaning forward

Sitting

Standing

Other

Body Movements

Walks around

Swiveling on chair

Change of position

Other

Head

Head tilt

Head pulled back

Head oriented away from camera

Sloped head posture

Straight head posture

Nodding

Shaking head

Other

Arms & Shoulders

Crosses arms

Arms behind back

Crosses arms behind back

Open arms while sitting

Shoulder movement

Other

Body touch

Touching head

Touching face

Touching hair

Touching body

Adjusting clothing

Other

Hands

Rubbing hands together

Empathetic gesture clapping

Circling hands

Waving

Hands folded

Symmetrical hand position

Talks while holding hand in front of mouth

Hands resting on table

Hands cling to something

Fidget or gesture w/ object

Interacting with device

Fast movements

Slow movements

Big gesture

Insulting gesture

Other

Fingers & Nails

Biting nails

Pointing

Rubbing fingers

Twisting fingers

Other

Visual Cue Inventory C

APPEARANCE

General Appearance

Groomed

Good personal cleanliness

Clothing

Colorful

Dark

Distinctive

Shiny

Very wrinkled

Shows a lot of skin

Businesslike

Non-athletic

Top

Pullover

Rolled-down sleeves

Loose-fitting

Bottom

Trousers

Loose-fitting

Hair

Hair covers part of face

Asymmetrical haircut

Plucked eyebrows

Untended hair

Undyed hair

Dark hair

Distinctive hair

Long hair

<u>If hair is rather long or long:</u>

Tied-up hair

<u>If beard is present:</u>

Groomed beard

Make up

No make up

<u>If make up is worn:</u>

Highlights

Other

Accessories

Glasses

Jewelry

Earrings

Necklace

Nose piercing

Tattoo

Headphones

MEDIA PROPERTIES

Visibility

Complete person and space around

Complete face

Eyes

Distance

Far away from camera

Far away from wall

Camera Position

Person in lower half of frame

Camera below head

Vertical picture frame

Lighting

Bright

Artificial

Even

No light halo

Face not lit by screen

Not overexposed

Light from behind

Image Quality

High resolution image

ENVIRONMENT

Others Present

Other person present

Background

Living room

Private location

Small room

Closely-cramped

Cluttered

Untidy

Wall Condition

Unfinished

Non-painted

No pattern

Window, Door, & Floor

No carpet

No door

Kitchen Equipment & Food

Fridge

Ice cube machine

Cookware & pots

Food

Furniture

Bed

Nightstand

Sofa

Chair

Desk

Table

Drawer

File cabinet

Garbage can

Wardrobe/closet

Shelve

Book shelve

Stereo stand

Crate

Coat rack

Tie rack

Objects – On Walls

Painting

Poster/photo

Flag

Mirror

Calendar

Clock

Objects – Entertaining

Book

Magazine

CD/record

Collection

Map

Game

Toy

Musical instrument

Objects – Useful

Stationary

Toiletry

Medication

Tool

Bag

Label

Lamp

Fan

Objects – Equipment

Electronic equipment

Athletic equipment

Weapon

Objects – Decoration

Decoration

Plant

Religious artifact

<u>If environment is decorated:</u>

Appropriate decoration

Objects – Textiles & Accessories in Background

Clothing item

Specialized clothing item

Jewelry

Pillow

Target:____________
Judge:_____________
Date:______________
Time:_____________

<u>Personality Ratings</u>

		Low				High
	Task Style	1	2	3	4	5
	Adaptation Style	1	2	3	4	5
	Achievement Style	1	2	3	4	5
	Teamwork Style	1	2	3	4	5
	Emotional Style	1	2	3	4	5
	Interaction Style	1	2	3	4	5

Visual Cue Inventory A

FACE

Facial Expression

	none	minimal		some	a lot	
Friendly	1	2	3	4	5	n.a.
Cheerful	1	2	3	4	5	n.a.
Interested	1	2	3	4	5	n.a.
Self-assured	1	2	3	4	5	n.a.
Timid	1	2	3	4	5	n.a.
Calm	1	2	3	4	5	n.a.
Surprised	1	2	3	4	5	n.a.
Skeptical	1	2	3	4	5	n.a.
Arrogant-amused	1	2	3	4	5	n.a.
Grumpy	1	2	3	4	5	n.a.
Indifferent	1	2	3	4	5	n.a.
Strong	1	2	3	4	5	n.a.
Diverse	1	2	3	4	5	n.a.
Rapidly changing	1	2	3	4	5	n.a.

( ) Other:______

Forehead

Wrinkled	1	2	3	4	5	n.a.

( ) Other:______

Look

At camera	1	2	3	4	5	n.a.
Away from camera sideways	1	2	3	4	5	n.a.
Up	1	2	3	4	5	n.a.
Down	1	2	3	4	5	n.a.

( ) Other:______

Eyes & Eyebrows

Rolling eyes	1	2	3	4	5	n.a.
Heavy blinking	1	2	3	4	5	n.a.
Wide open eyes	1	2	3	4	5	n.a.
Closed eyes	1	2	3	4	5	n.a.
Narrow eyes	1	2	3	4	5	n.a.
Tinkles with eyelashes	1	2	3	4	5	n.a.
Furrowed eyebrows	1	2	3	4	5	n.a.
Raised eyebrows	1	2	3	4	5	n.a.

( ) Other:______

Lips

Laughing____	1	2	3	4	5	n.a.
Smiling____	1	2	3	4	5	n.a.
Biting lips____	1	2	3	4	5	n.a.
Licking lips____	1	2	3	4	5	n.a.
Pressed lips____	1	2	3	4	5	n.a.
Plops lips____	1	2	3	4	5	n.a.
Duckface____	1	2	3	4	5	n.a.
Pout____	1	2	3	4	5	n.a.

( ) Other:____

Mouth & Throat

Wide open mouth____	1	2	3	4	5	n.a.
Pauses____	1	2	3	4	5	n.a.
Swallowing hard____	1	2	3	4	5	n.a.
Fast mouth movements____	1	2	3	4	5	n.a.

( ) Other:____

IMAGE QUALITY

Camera shaking____	1	2	3	4	5	n.a.

( ) Other:____

BODY

Legs

Shaking legs____	1	2	3	4	5	n.a.
Legs crossed____	1	2	3	4	5	n.a.
Legs open____	1	2	3	4	5	n.a.

( ) Other:____

Visual Cue Inventory B

BODY

none minimal some a lot

Body Posture

Straight posture	1	2	3	4	5 n.a.
Slouched posture	1	2	3	4	5 n.a.
Relaxed posture	1	2	3	4	5 n.a.
Maintained posture	1	2	3	4	5 n.a.
Open posture	1	2	3	4	5 n.a.
Stiff posture	1	2	3	4	5 n.a.
Body tilted	1	2	3	4	5 n.a.
Body turned away	1	2	3	4	5 n.a.
Parallel body orientation	1	2	3	4	5 n.a.
Leaning backward	1	2	3	4	5 n.a.
Leaning forward	1	2	3	4	5 n.a.
Sitting	1	2	3	4	5 n.a.
Standing	1	2	3	4	5 n.a.

( ) Other:

Body Movements

Walks around	1	2	3	4	5 n.a.
Swiveling on chair	1	2	3	4	5 n.a.
Change of position	1	2	3	4	5 n.a.

( ) Other:

Head

Head tilt	1	2	3	4	5 n.a.
Head pulled back	1	2	3	4	5 n.a.
Head oriented away from camera	1	2	3	4	5 n.a.
Sloped head posture	1	2	3	4	5 n.a.
Straight head posture	1	2	3	4	5 n.a.
Nodding	1	2	3	4	5 n.a.
Shaking head	1	2	3	4	5 n.a.

( ) Other:

Arms & Shoulders

Crosses arms	1	2	3	4	5 n.a.
Arms behind back	1	2	3	4	5 n.a.
Crosses arms behind back	1	2	3	4	5 n.a.
Open arms while sitting	1	2	3	4	5 n.a.
Shoulder movement	1	2	3	4	5 n.a.

( ) Other:

Body touch

Touching head	1	2	3	4	5 n.a.
Touching face	1	2	3	4	5 n.a.
Touching hair	1	2	3	4	5 n.a.
Touching body	1	2	3	4	5 n.a.
Adjusting clothing	1	2	3	4	5 n.a.

( ) Other:

Hands

Rubbing hands together	1	2	3	4	5	n.a.
Empathetic gesture clapping	1	2	3	4	5	n.a.
Circling hands	1	2	3	4	5	n.a.
Waving	1	2	3	4	5	n.a.
Hands folded	1	2	3	4	5	n.a.
Symmetrical hand position	1	2	3	4	5	n.a.
Talks while holding hand in front of mouth	1	2	3	4	5	n.a.
Hands resting on table	1	2	3	4	5	n.a.
Hands cling to something	1	2	3	4	5	n.a.
Fidget or gesture w/ object	1	2	3	4	5	n.a.
Interacting with device	1	2	3	4	5	n.a.
Fast movements	1	2	3	4	5	n.a.
Slow movements	1	2	3	4	5	n.a.
Big gesture	1	2	3	4	5	n.a.
Insulting gesture	1	2	3	4	5	n.a.

( ) Other:

Fingers & Nails

Biting nails	1	2	3	4	5	n.a.
Pointing	1	2	3	4	5	n.a.
Rubbing fingers	1	2	3	4	5	n.a.
Twisting fingers	1	2	3	4	5	n.a.

( ) Other:

Visual Cue Inventory C

APPEARANCE

General Appearance

Ungroomed	1	2	3	4	Groomed	5	n.a.
Poor personal cleanliness	1	2	3	4	Good personal cleanliness	5	n.a.

Clothing

Neutral	1	2	3	4	Colorful	5	n.a.
Dark	1	2	3	4	Light	5	n.a.
Plain	1	2	3	4	Distinctive	5	n.a.
Dull	1	2	3	4	Shiny	5	n.a.
Wrinkle-free	1	2	3	4	Very wrinkled	5	n.a.
Shows no skin	1	2	3	4	Shows a lot of skin	5	n.a.
Casual	1	2	3	4	Businesslike	5	n.a.
Athletic	1	2	3	4	Non-athletic	5	n.a.

Top

( ) Tank Top ( ) T-Shirt ( ) Button-up shirt/blouse ( ) Pullover
( ) Other:___________ ( ) n.a.

Rolled-up sleeves	1			4	Rolled-down sleeves	5	n.a.
Close-fitting	1	2	3	4	Loose-fitting	5	n.a.

Bottom

( ) Trousers ( ) Other:________________________ ( ) n.a.

Close-fitting	1	2	3	4	Loose-fitting	5	n.a.

Hair

Hair does not cover face	1			4	Hair covers part of face	5	n.a.
Symmetrical haircut	1			4	Asymmetrical haircut	5	n.a.
Natural eyebrows	1			4	Plucked eyebrows	5	n.a.
Tended hair	1	2	3	4	Untended hair	5	n.a.
Dyed hair	1			4	Undyed hair	5	n.a.
Light hair	1	2	3	4	Dark hair	5	n.a.
Plain hair	1	2	3	4	Distinctive hair	5	n.a.
Short hair	1	2	3	4	Long hair	5	n.a.
<u>If hair is rather long or long:</u>							
Open hair	1	2	3	4	Tied-up hair	5	n.a.
<u>If beard is present:</u>							
Groomed beard	1	2	3	4	Ungroomed beard	5	n.a.

Make up

Heavy make up	1	2	3	4	No make up	5	n.a.

<u>If make up is worn:</u>
( ) Made-up eyes ( ) Rouged lips ( ) Highlights ( ) Other:________________

Accessories

Glasses	1			4 No glasses	5 n.a.
Jewelry	1			4 No jewelry	5 n.a.
Earrings	1			4 No earrings	5 n.a.
Necklace	1			4 No necklace	5 n.a.
Nose piercing	1			4 No nose piercing	5 n.a.
Tattoo	1			4 No tattoo	5 n.a.

( ) Headphones ( ) Headset ( ) No headphones/-set ( ) n.a.

( ) Other:____

MEDIA PROPERTIES

Visibility

( ) Complete person and space around ( ) Complete person ( ) Upper body
( ) Only head/face ( ) No parts of person ( ) n.a.

( ) Complete face ( ) Only parts of face ( ) No parts of face ( ) n.a.

( ) Eyes ( ) Hands ( ) Gestures ( ) n.a.

Distance

Far away from camera	1	2	3	4 Close to camera	5 n.a.
Far away from wall	1	2	3	4 Close to wall	5 n.a.

Camera Position

( ) Person in lower half of frame ( ) Person in center ( ) Other:___________ ( ) n.a.

( ) Camera below head ( ) Camera level with head ( ) Camera above head ( ) n.a.

( ) Vertical picture frame ( ) Horizontal picture frame ( ) Other:___________ ( ) n.a.

Lighting

Dark	1	2	3	4 Bright	5 n.a.
Natural	1	2	3	4 Artificial	5 n.a.
Uneven	1	2	3	4 Even	5 n.a.
Light halo	1			4 No light halo	5 n.a.
Face strongly lit by screen	1	2	3	4 Face not lit by screen	5 n.a.
Overexposed	1			4 Not overexposed	5 n.a.

( ) Light from behind ( ) Light from left side ( ) Light from right side
( ) Other: _____________________ ( ) n.a.

Image Quality

High resolution image	1	2	3	4 Low resolution image	5 n.a.

ENVIRONMENT

Others Present

( ) Other person present ( ) Pet present ( ) Nobody else present ( ) n.a.

Background

( ) Living room ( ) Kitchen ( ) Bedroom ( ) Other: _______________ ( ) n.a.

Private location	1			4	Public location	5	n.a.
Big room	1	2	3	4	Small room	5	n.a.
Open-spacious	1	2	3	4	Closely-cramped	5	n.a.
Plain	1	2	3	4	Cluttered	5	n.a.
Tidy	1	2	3	4	Untidy	5	n.a.

Wall Condition

Finished	1	4	Unfinished	5	n.a.
Painted	1	4	Non-painted	5	n.a.
Pattern	1	4	No pattern	5	n.a.

Window, Door, & Floor

( ) Blinds ( ) Shutters ( ) Curtains ( ) No window covering ( ) No window ( ) n.a.

Carpet	1	4	No carpet	5	n.a.
Door	1	4	No door		

( ) Other:____

Kitchen Equipment & Food

Fridge	1	4	No fridge
Ice cube machine	1	4	No ice cube machine
Cookware & pots	1	4	No cookware & pots
Food	1	4	No food

( ) Other:____

Furniture

Bed	1	4	No bed
Nightstand	1	4	No nightstand
Sofa	1	4	No sofa
Chair	1	4	No chair
Desk	1	4	No desk
Table	1	4	No table
Drawer	1	4	No drawer
File cabinet	1	4	No file cabinet
Garbage can	1	4	No garbage can
Wardrobe/closet	1	4	No wardrobe/closet
Shelve	1	4	No shelve
Book shelve	1	4	No book shelve
Stereo stand	1	4	No stereo stand
Crate	1	4	No crate
Coat rack	1	4	No coat rack
Tie rack	1	4	No tie rack

( ) Other:____

Objects – On Walls

Painting	1	4	No painting
Poster/photo	1	4	No poster/photo
Flag	1	4	No flag
Mirror	1	4	No mirror
Calendar	1	4	No calendar
Clock	1	4	No clock

( ) Other:______________________________

Objects – Entertaining

Book	1	4	No book
Magazine	1	4	No magazine
CD/record	1	4	No CD/record
Collection	1	4	No collection
Map	1	4	No map
Game	1	4	No game
Toy	1	4	No toy
Musical instrument	1	4	No musical instrument

( ) Other:______________________________

Objects – Useful

Stationary	1	4	No stationary
Toiletry	1	4	No toiletry
Medication	1	4	No medication
Tool	1	4	No tool
Bag	1	4	No bag
Label	1	4	No label
Lamp	1	4	No lamp
Fan	1	4	No fan

( ) Other:______________________________

Objects – Equipment

Electronic equipment	1	4	No electronic equipment
Athletic equipment	1	4	No athletic equipment
Weapon	1	4	No weapon

( ) Other:______________________________

Objects – Decoration

Decoration	1	4	No decoration
Plant	1	4	No plant
Religious artifact	1	4	No religious artifact

( ) Other:______________________________

<u>If environment is decorated:</u>

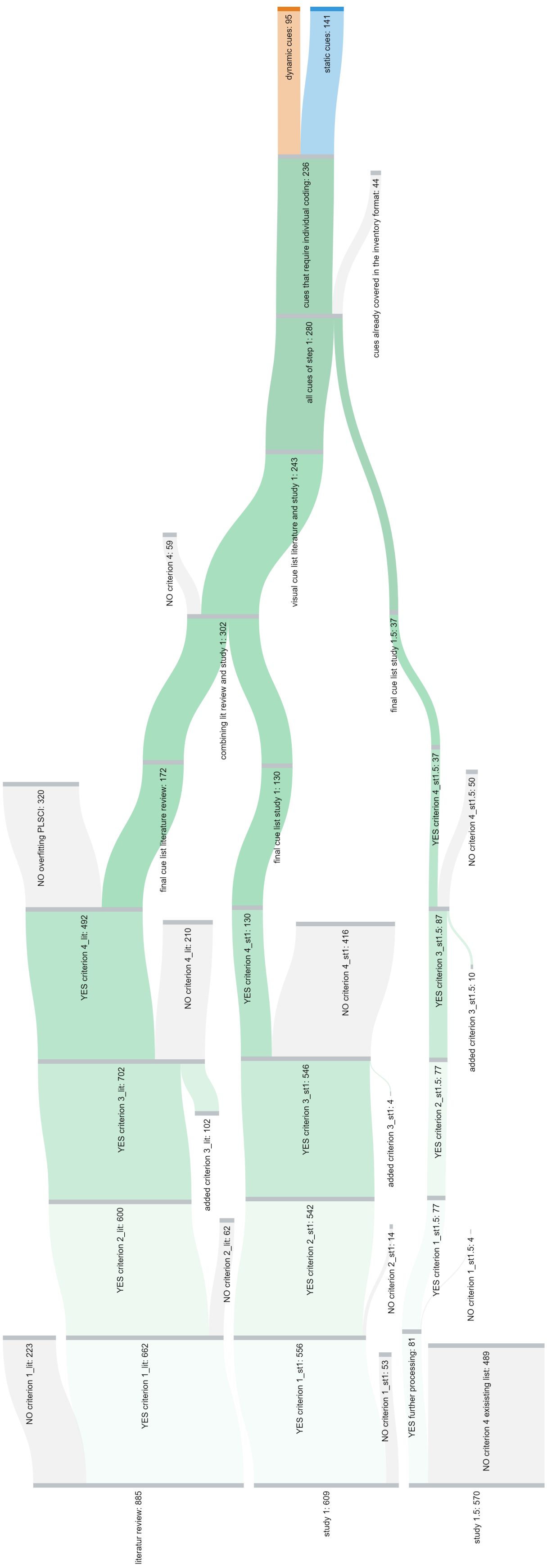
Appropriate decoration	1	4	Inappropriate decoration

Objects – Textiles & Accessories in Background

Clothing item	1	4	No clothing item
Specialized clothing item	1	4	No specialized clothing item
Jewelry	1	4	No jewelry
Pillow	1	4	No pillow

( ) Other:______________________________

XI – Cue Processing Flow



XII – Job Advertisement from dataset 2

Job Advertisement: Manager Trainee in a Grocery Chain

Job Description

Company EXCELLENCE is one of America's favorite grocery stores. Their focus on customer service and product quality has driven the company to grow in the past few years. Currently, EXCELLENCE is hiring for a full-time manager trainee. The manager trainee will be responsible for merchandising products, monitoring inventory and keeping the store looking its best. This position has an opportunity to grow to a store manager, or transition to the corporate with different departments such as purchasing, logistics, or customer care.

Responsibilities

- Assists management with developing and implementing action plans to drive results.
- Establishes and communicates job responsibilities and performance expectations to the team to assure mutual understanding of desired results; resolves internal or external barriers that prohibit successful goal achievement.
- Provides product feedback to the management, including making recommendations regarding new items to carry or those that should be discontinued.
- Ensures an appropriate resolution of operational customer concerns in management's absence.
- Ensures a safe environment for employees, customers and vendors.
- Ensures the quality and freshness of products to maximize sales.
- Identifies cost-saving opportunities and potential process improvements.

Knowledge/Skills/Abilities

- Provides prompt and courteous customer service.
- Ability to operate a cash register efficiently and accurately under pressure.
- Excellent written and verbal communication.
- Gives attention to detail and follows instructions.
- Ability to work both independently and within a team environment.
- Ability to supervise store personnel in the store manager's absence to ensure the timely and effective completion of work assignments.
- Understands and applies management principles concerning budgeting, personnel costs, and overtime expenses.
- Ability to develop rapport, trust, and open communication that enhances the growth and job performance of direct reports.
- Ability to prioritize and work under strict deadlines.

XIII – Participant Instructions from dataset 2

This study aims to gain insights into video interviews. A web camera is required for completing this HIT. In the following section, you will record and upload your video responses to six interview questions for the trainee position. Please provide **work-relevant examples** and different work-relevant examples for each interview question. Please imagine yourself as actually applying for the position, and make sure your physical appearance is appropriate for a professional interview setting (e.g., do not wear pajamas or cover yourself with a blanket).

Please make sure to **turn on your camera** and **test it** before taking the assessment, so the visual and audio recording can be as clear as possible. All the recordings will be used for research purposes only, and we will not share any of your information outside of the research team without your consent. Your responses will be rated by an artificial intelligent tool and human raters.

Please be reminded that the best 10 candidates (i.e. participants) of this study will be **rewarded a $10 bonus**. Responses that have one or more of the following elements will disqualify you for receiving any type of compensation for completing the assessment:

- Examples are not relevant to the workplace (e.g., communication with a pet, an incident with a personal friend that does not involve any workplace content)
- Lack of effort (e.g., answers are shorter than 90 seconds)
- Poor audio quality (e.g., answers cannot be heard clearly)
- Inappropriate self-presentation (e.g., wearing pajamas, laying in bed)
- Inattentive responding to survey items (e.g., going through the survey extremely quickly)

Lastly, if you ran out of time before completing the assessment, please submit the HIT then continue taking the assessment. If you have any trouble submitting your videos, please contact us at mturk@cut-e.com.

XIV – Interview Questions from dataset 2

1. Give me an example of a time when you were able to maintain a high level of focus when working under pressure (e.g. tight deadline, heavy workload, etc.).
    a. What was the situation?
    b. How did you cope with this situation?
    c. How did you react to the pressure?
    d. What were the results of your actions?

2. Give me an example of a time when you faced a particularly stressful or uncomfortable interaction at work.
    a. What was the situation?
    b. What about the interaction was stressful or uncomfortable?
    c. How did you cope with this situation?
    d. What were the results of your actions?

3. Describe a time when you had to follow detailed instructions to complete a task.
    a. What was the situation? What procedures did you need to follow?
    b. How did you ensure that each step was done accurately?
    c. Did you make any mistakes along the way? If so, how did you find the mistake and correct the issue?
    d. To what degree did you follow all of the instructions? Did you deviate from them at all? If so, why?
    e. What was the end result?

4. Describe a time when you had to perform a task that required careful attention to detail and quality.

   a. What steps did you take to maintain the organizations or your own quality standards?

   b. Who would be affected if the quality of your work did not meet standards?

   c. Describe the types of details you were required to attend to.

   d. How did you ensure that important details were not missed?

   e. What was the final outcome?

5. Describe a situation that required you to work with others to help identify and meet a customer’s request or needs.

   a. What was the situation?

   b. What process did you use to determine the most effective response to the client or customer?

   c. What did you do to ensure the request was accurately met?

   d. What was the result of your actions?

6. Suppose a customer or client approaches you and cannot find a product he or she is looking for. How would you handle the situation?

   a. How would you complete his or her request?

   b. What would you do to gather additional information?

   c. What would you do to ensure all requests were accurately met?

   d. How would you know if the customer was satisfied?

XV – Details overview of technical feature descriptive statistics

	N	M	SD	Min	Max	Skew	Kurtosis
Q1_WordCount	98	312.58	92.02	98.00	436.00	-0.60	-0.89
Q2_WordCount	97	307.23	90.56	87.00	437.00	-0.50	-0.98
Q3_WordCount	98	315.93	86.65	57.00	432.00	-0.80	-0.16
Q4_WordCount	99	317.66	84.03	77.00	426.00	-0.84	-0.03
Q5_WordCount	99	300.38	86.87	80.00	437.00	-0.44	-0.68
Q6_WordCount	99	309.80	85.47	37.00	429.00	-0.67	0.07
Q1_CharacterCount	98	1267.79	378.88	370.00	1661.00	-0.58	-1.00
Q2_CharacterCount	97	1243.06	364.95	393.00	1677.00	-0.50	-1.11
Q3_CharacterCount	98	1294.06	353.59	268.00	1671.00	-0.80	-0.22
Q4_CharacterCount	99	1315.60	345.18	309.00	1669.00	-0.86	-0.19
Q5_CharacterCount	99	1234.69	359.45	352.00	1680.00	-0.47	-0.76
Q6_CharacterCount	99	1248.17	348.22	162.00	1657.00	-0.67	-0.04
Q1_Confidence	98	77.71	10.29	40.00	94.00	-0.79	0.42
Q2_Confidence	97	76.89	10.15	56.00	94.00	-0.35	-0.89
Q3_Confidence	98	76.50	9.17	53.00	91.00	-0.61	-0.40
Q4_Confidence	99	75.67	10.04	50.00	90.00	-0.59	-0.64
Q5_Confidence	99	76.82	9.50	53.00	94.00	-0.35	-0.92
Q6_Confidence	99	76.53	10.18	43.00	94.00	-0.72	0.13
Q1_Audio Quality	92	3.99	1.52	0.00	5.00	-1.68	1.64
Q2_Audio Quality	92	4.02	1.54	0.00	5.00	-1.66	1.51
Q3_Audio Quality	92	4.16	1.51	0.00	5.00	-1.94	2.44
Q4_Audio Quality	92	4.18	1.42	0.00	5.00	-1.90	2.57
Q5_Audio Quality	92	4.10	1.49	0.00	5.00	-1.73	1.85
Q6_Audio Quality	92	4.10	1.50	0.00	5.00	-1.81	2.10
Q1_Video Quality	92	4.34	1.44	0.00	5.00	-2.36	4.21
Q2_Video Quality	92	4.32	1.48	0.00	5.00	-2.29	3.77
Q3_Video Quality	92	4.37	1.44	0.00	5.00	-2.46	4.59
Q4_Video Quality	92	4.45	1.33	0.00	5.00	-2.68	5.88
Q5_Video Quality	92	4.39	1.41	0.00	5.00	-2.48	4.73
Q6_Video Quality	92	4.32	1.48	0.00	5.00	-2.29	3.76
WordCount_av	97	311.82	76.62	89.33	425.50	-0.58	-0.38
CharCount_av	97	1272.08	315.29	400.67	1653.67	-0.58	-0.55
Confidence_av	97	76.69	9.22	55.33	89.83	-0.50	-0.87
AudioQual_av	92	4.09	1.35	0.00	5.00	-1.67	1.90
VideoQual_av	92	4.36	1.26	0.00	5.00	-2.40	4.99

XVI – Regression Models

The regression models are displayed in the following order:

- Per table, both self-rating and corresponding observer ratings for each trait are displayed next to each other.
- The first 5 tables display the regression models using the IPIP datapoints (models 1 – 5 and 11 – 15)
- The next 5 tables display the regression models using the ADEPT datapoints (models 10 – 14 and 16 – 20)
- The last 5 tables display the regression models using the ADEPT NLC datapoints (models 21 – 25). Although duplication, the ADEPT self-rating regression is included in those overviews again to better see the (miss-) alignment between the two regression models and the variables that are used for them.

Regression Models IPIP

	IPIP self-rating Openness					IPIP observer rating Openness				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	10.59	0	8.24 – 12.95	-0.17 – 0.17	**<0.001**	13.4	0	10.55 – 16.24	-0.18 – 0.18	**<0.001**
Pullover	0.29	0.13	-0.11 – 0.68	-0.05 – 0.31	0.154					
Hair cover face	0.59	0.29	0.21 – 0.97	0.10 – 0.47	**0.003**					
Earrings	1.03	0.31	0.45 – 1.62	0.13 – 0.49	**0.001**					
chair	-0.31	-0.13	-0.73 – 0.11	-0.31 – 0.05	0.146					
Map	1.17	0.21	0.17 – 2.18	0.03 – 0.39	**0.023**	0.49	0.1	-0.68 – 1.66	-0.14 – 0.34	0.405
Decoration	-0.41	-0.16	-0.92 – 0.10	-0.36 – 0.04	0.112					
Plant	-0.83	-0.15	-1.93 – 0.27	-0.34 – 0.05	0.136					
Heavy blinking	-0.87	-0.25	-1.67 – -0.07	-0.47 – -0.02	**0.033**					
Closed eyes	-0.64	-0.15	-1.67 – 0.39	-0.40 – 0.09	0.222					
Narrow eyes	-0.13	-0.04	-0.76 – 0.51	-0.24 – 0.16	0.692					
Pressed lips	-0.12	-0.04	-0.80 – 0.56	-0.24 – 0.17	0.73					
Business-like clothes						0.23	0.1	-0.19 – 0.64	-0.09 – 0.29	0.284
Only parts of face						-1.03	-0.21	-1.94 – -0.13	-0.40 – -0.03	**0.026**
Book shelve						0.12	0.05	-0.66 – 0.91	-0.26 – 0.36	0.753
Calendar						0.64	0.13	-0.45 – 1.74	-0.09 – 0.36	0.244
Book						0.26	0.13	-0.39 – 0.92	-0.19 – 0.45	0.428
Stationary						-0.1	-0.02	-1.00 – 0.80	-0.25 – 0.20	0.829
Athletic equipment						-0.78	-0.11	-2.16 – 0.61	-0.32 – 0.09	0.266
Friendly expressions						0.24	0.09	-0.32 – 0.80	-0.12 – 0.29	0.397
Timid expressions						-0.46	-0.12	-1.25 – 0.34	-0.33 – 0.09	0.259
Strong expressions						0.7	0.25	0.16 – 1.24	0.06 – 0.45	**0.012**
Pausing						-0.8	-0.26	-1.43 – -0.17	-0.46 – -0.05	**0.014**
Observations	84					88				
R2 / R2 adjusted	0.493 / 0.415					0.369 / 0.268				
F-statistic (*p*)	6.36 (< **0.001**)					3.65 **(< 0.001**)				

	IPIP self-rating Conscientiousness					IPIP observer-rating Conscientiousness				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	12.5	0	10.88 – 14.12	-0.21 – 0.21	**<0.001**	3.12	0	-8.02 – 14.25	-0.21 – 0.21	0.572
Long hair	-0.39	-0.26	-0.71 – -0.06	-0.47 – -0.04	**0.02**					
Nose piercing	0.42	0.11	-0.49 – 1.33	-0.12 – 0.33	0.36					
chair	-0.25	-0.14	-0.62 – 0.12	-0.36 – 0.07	0.186					
Bag	0.55	0.17	-0.20 – 1.31	-0.06 – 0.40	0.149					
Laughing	-0.53	-0.17	-1.23 – 0.16	-0.38 – 0.05	0.127					
Good personal cleanliness						0.3	0.14	-0.54 – 1.13	-0.25 – 0.52	0.475
Colorful clothes						0.5	0.27	0.00 – 0.99	0.00 – 0.53	**0.049**
Distinctive hair						-0.15	-0.08	-0.69 – 0.38	-0.37 – 0.20	0.563
Only parts of face						0.06	0.01	-1.90 – 2.03	-0.41 – 0.43	0.947
Overexposed						0.08	0.05	-0.30 – 0.45	-0.22 – 0.32	0.682
Bedroom						-0.37	-0.25	-0.75 – 0.02	-0.51 – 0.01	0.059
cookware & pots						0.21	0.05	-0.97 – 1.40	-0.21 – 0.30	0.716
Shelve						0.01	0	-0.52 – 0.54	-0.31 – 0.32	0.981
Toy						0.49	0.15	-0.47 – 1.45	-0.14 – 0.43	0.304
Stationary						1	0.21	-0.17 – 2.16	-0.04 – 0.46	0.092
Athletic equipment						-0.03	-0.01	-1.46 – 1.41	-0.31 – 0.30	0.97
Decoration						0.31	0.16	-0.15 – 0.78	-0.07 – 0.38	0.176
Pillow						0.09	0.04	-0.59 – 0.76	-0.29 – 0.38	0.798
Timid expressions						0.43	0.14	-0.41 – 1.27	-0.13 – 0.41	0.3
Strong expressions						0.62	0.26	0.04 – 1.21	0.02 – 0.51	**0.037**
Looking at camera						0.79	0.32	0.04 – 1.54	0.01 – 0.63	**0.041**
Looking down						0.04	0.02	-0.53 – 0.61	-0.26 – 0.30	0.877
Pausing						-0.3	-0.11	-1.15 – 0.55	-0.42 – 0.20	0.48
Straight posture						0.53	0.24	-0.08 – 1.13	-0.04 – 0.52	0.088
Sitting						0.98	0.27	-0.78 – 2.74	-0.21 – 0.74	0.266
Observations	78					52				
R2 / R2 adjusted	0.210 / 0.155					0.667 / 0.452				
F-statistic (*p*)	3.84 (**0.004**)					3.11 (**0.002**)				

	IPIP self-rating Extraversion					IPIP observer-rating Extraversion				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	13.02	0	9.38 – 16.66	-0.22 – 0.22	**<0.001**	15.67	0	12.64 – 18.71	-0.21 – 0.21	**<0.001**
Long hair	-0.55	-0.3	-1.04 – -0.06	-0.56 – -0.03	**0.028**	-0.36	-0.21	-0.76 – 0.04	-0.45 – 0.03	0.08
Camera level with head	0.22	0.14	-0.19 – 0.62	-0.12 – 0.41	0.286					
Dark light	0.5	0.18	-0.21 – 1.22	-0.08 – 0.43	0.164					
Overexposed	-0.16	-0.1	-0.60 – 0.28	-0.37 – 0.17	0.465					
Window	0.5	0.26	0.02 – 0.97	0.01 – 0.51	**0.041**	-0.26	-0.13	-0.69 – 0.18	-0.36 – 0.09	0.243
Wardrobe/closet	-0.15	-0.06	-0.80 – 0.51	-0.30 – 0.19	0.656					
Athletic equipment	-0.47	-0.11	-1.62 – 0.68	-0.36 – 0.15	0.414					
Sceptical expressions	-0.14	-0.04	-1.41 – 1.13	-0.36 – 0.28	0.824					
Wrinkled forehead	-0.17	-0.06	-1.21 – 0.86	-0.45 – 0.32	0.741					
Rolling eyes	-0.12	-0.04	-1.42 – 1.17	-0.51 – 0.42	0.848					
Wide open eyes	0.38	0.16	-0.56 – 1.31	-0.23 – 0.54	0.421					
Raised eyebrows	-0.56	-0.22	-1.47 – 0.34	-0.58 – 0.14	0.218					
Pout	-1.86	-0.23	-4.53 – 0.81	-0.55 – 0.10	0.168					
Pausing	-0.03	-0.01	-1.14 – 1.08	-0.35 – 0.33	0.952					
Swallowing hard	0.19	0.06	-1.26 – 1.64	-0.39 – 0.51	0.794					
Sloped head posture	0.08	0.02	-1.25 – 1.41	-0.32 – 0.36	0.906					
Shaking head	0.49	0.17	-0.53 – 1.50	-0.19 – 0.52	0.344					
Colorful clothes						0.18	0.08	-0.30 – 0.67	-0.14 – 0.31	0.454
Pullover						-0.18	-0.1	-0.61 – 0.26	-0.36 – 0.15	0.419
Even light						-0.26	-0.13	-0.70 – 0.19	-0.35 – 0.10	0.256
cookware & pots						0.85	0.15	-0.43 – 2.14	-0.08 – 0.38	0.19
Decoration						0.28	0.16	-0.12 – 0.69	-0.07 – 0.39	0.17
Timid expressions						-1.16	-0.31	-1.96 – -0.35	-0.53 – -0.10	**0.005**
Strong expressions						0.44	0.16	-0.31 – 1.19	-0.11 – 0.44	0.243
Diverse facial expressions						-0.86	-0.3	-2.45 – 0.73	-0.85 – 0.25	0.281
Rapidly changing facial expressions						0.74	0.27	-0.78 – 2.27	-0.28 – 0.81	0.334
Observations	70					69				
R2 / R2 adjusted	0.345 / 0.130					0.388 / 0.269				
F-statistic (*p*)	1.61 (0.096)					3.28 (**0.002**)				

	IPIP self-rating Agreeableness					IPIP observer rating Agreeableness				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	9.5	0	6.99 – 12.02	-0.18 – 0.18	**<0.001**	16.32	0	8.21 – 24.43	-0.20 – 0.20	**0.001**
Necklace	0.76	0.24	0.16 – 1.36	0.05 – 0.43	**0.013**					
Dark light	0.6	0.27	0.16 – 1.03	0.07 – 0.47	**0.008**					
Even light	0.52	0.29	0.16 – 0.87	0.09 – 0.49	**0.005**					
Shelve	-0.27	-0.15	-0.60 – 0.06	-0.35 – 0.04	0.113					
Surprised expressions	0.3	0.08	-0.51 – 1.11	-0.14 – 0.31	0.461					
Licking lips	-0.44	-0.14	-1.19 – 0.31	-0.37 – 0.10	0.247					
Plopping lips	-0.29	-0.1	-1.03 – 0.46	-0.37 – 0.16	0.445	-0.98	-0.33	-2.85 – 0.88	-0.97 – 0.30	0.271
Wide open mouth	-0.14	-0.06	-0.64 – 0.36	-0.27 – 0.16	0.586					
Distinctive clothes						-0.06	-0.03	-0.80 – 0.68	-0.34 – 0.29	0.856
Business-like clothes						0.35	0.19	-0.28 – 0.97	-0.15 – 0.52	0.25
Loose-fitting top						-1.6	-0.77	-2.39 – -0.81	-1.16 – -0.39	**0.001**
Distinctive hair						0.29	0.15	-0.41 – 0.98	-0.21 – 0.51	0.381
Big distance to wall						0.22	0.12	-0.43 – 0.86	-0.24 – 0.47	0.477
Overexposed						0.89	0.54	0.03 – 1.75	0.02 – 1.07	**0.045**
Tidy room						0.29	0.16	-0.30 – 0.88	-0.17 – 0.50	0.308
Stereo stand						0.47	0.11	-3.91 – 4.85	-0.93 – 1.15	0.817
Athletic equipment						0.52	0.17	-2.71 – 3.75	-0.90 – 1.24	0.73
Pillow						-1.67	-0.84	-2.93 – -0.41	-1.48 – -0.21	**0.014**
Timid expressions						-1.14	-0.41	-2.34 – 0.06	-0.83 – 0.02	0.06
Rolling eyes						1.74	0.82	0.49 – 3.00	0.23 – 1.41	**0.011**
Furrowed eyebrows						-0.14	-0.07	-0.96 – 0.68	-0.45 – 0.32	0.707
Biting lips						-0.02	0	-3.08 – 3.04	-0.80 – 0.79	0.99
Pressed lips						-0.36	-0.16	-1.72 – 1.00	-0.76 – 0.44	0.572
Duckface						-9.22	-1.54	-21.84 – 3.39	-3.66 – 0.57	0.136
Pout						8.83	1.47	-2.31 – 19.97	-0.38 – 3.33	0.109
Pausing						-0.02	-0.01	-1.39 – 1.35	-0.55 – 0.54	0.977
Straight posture						-0.08	-0.04	-1.24 – 1.09	-0.58 – 0.51	0.886
Slouched posture						0.1	0.05	-1.16 – 1.37	-0.59 – 0.69	0.859
Body tilted						0.66	0.28	-0.51 – 1.83	-0.22 – 0.78	0.24
Body turned away						-1.27	-0.35	-3.32 – 0.78	-0.90 – 0.21	0.199
Leaning forward						-0.6	-0.25	-1.78 – 0.57	-0.75 – 0.24	0.282
Head oriented away from camera						-1.99	-0.75	-3.80 – -0.17	-1.44 – -0.07	**0.035**
Sloped head posture						0.2	0.06	-1.09 – 1.49	-0.30 – 0.41	0.737
Shoulder movement						-1.47	-0.64	-2.69 – -0.26	-1.16 – -0.11	**0.022**
Touching head						0.2	0.06	-1.40 – 1.80	-0.39 – 0.51	0.788
Touching hair						3.96	1.24	0.81 – 7.11	0.25 – 2.23	**0.018**
Touching body						0.59	0.16	-1.27 – 2.45	-0.34 – 0.65	0.502
Observations	96					42				
R2 / R2 adjusted	0.273 / 0.206					0.907 / 0.654				
F-statistic	4.08 (< **0.001**)					3.59 (**0.015**)				

	IPIP self-rating Neuroticism					IPIP self-rating Neuroticism				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	18.57	0	14.64 – 22.51	-0.19 – 0.19	**<0.001**	11.68	0	10.04 – 13.32	-0.24 – 0.24	**<0.001**
Dyed hair	0.37	0.18	-0.10 – 0.84	-0.05 – 0.41	0.117					
Vertical picture frame	-0.69	-0.2	-1.36 – -0.01	-0.40 – -0.00	**0.046**					
Dark light	0.67	0.25	0.10 – 1.25	0.04 – 0.47	**0.022**					
Even light	-0.22	-0.11	-0.65 – 0.22	-0.32 – 0.11	0.322	0.15	0.19	-0.08 – 0.39	-0.10 – 0.49	0.19
Face lit by screen	-0.19	-0.08	-0.69 – 0.31	-0.30 – 0.14	0.458					
Timid expressions	-0.23	-0.07	-1.07 – 0.61	-0.30 – 0.17	0.578					
Calm expressions	-0.49	-0.17	-1.18 – 0.20	-0.41 – 0.07	0.162					
Wrinkled forehead	-0.29	-0.12	-0.89 – 0.31	-0.38 – 0.13	0.331					
Looking at camera	-0.94	-0.35	-1.54 – -0.34	-0.57 – -0.12	**0.003**					
Raised eyebrows	-0.46	-0.19	-1.13 – 0.21	-0.47 – 0.09	0.175					
Pout	-1.27	-0.18	-3.12 – 0.58	-0.44 – 0.08	0.176					
Fast mouth movements	0.13	0.06	-0.46 – 0.73	-0.20 – 0.32	0.655					
Slouched posture	-0.25	-0.12	-0.75 – 0.26	-0.36 – 0.12	0.33	0.07	0.09	-0.16 – 0.30	-0.21 – 0.39	0.539
Straight head posture	0.17	0.06	-0.53 – 0.86	-0.19 – 0.31	0.633					
Shaking head	-0.23	-0.09	-0.95 – 0.49	-0.37 – 0.19	0.524					
Good personal cleanliness						-0.05	-0.05	-0.35 – 0.25	-0.39 – 0.28	0.742
Hair cover face						-0.08	-0.14	-0.24 – 0.08	-0.42 – 0.14	0.328
Asymmetrical haircut						-0.03	-0.05	-0.20 – 0.14	-0.34 – 0.25	0.755
Tended hair						0.21	0.27	-0.03 – 0.44	-0.04 – 0.58	0.09
Heavy make up						0.32	0.26	-0.00 – 0.63	-0.00 – 0.51	0.051
Light from right side						-0.07	-0.11	-0.25 – 0.11	-0.40 – 0.18	0.45
Kitchen						0.01	0.01	-0.26 – 0.28	-0.26 – 0.27	0.963
Shelve						-0.12	-0.19	-0.30 – 0.06	-0.47 – 0.09	0.186
Surprised expressions						-0.22	-0.15	-0.64 – 0.21	-0.44 – 0.14	0.308
Head tilt						0.04	0.05	-0.23 – 0.32	-0.25 – 0.34	0.744
Observations	69					54				
R2 / R2 adjusted	0.529 / 0.396					0.427 / 0.260				
F-statistic (*p*)	3.97 (< **0.001**)					2.55 (**0.013**)				

Regression Models ADEPT

	ADEPT self-rating Openness					ADEPT observer rating Openness				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	-0.33	0	-0.56 – -0.11	-0.19 – 0.19	**0.004**	0.52	0	-2.58 – 3.61	-0.20 – 0.20	0.738
Electronic equipment	0.08	0.25	0.02 – 0.14	0.06 – 0.45	**0.013**					
desk	0.06	0.12	-0.05 – 0.17	-0.10 – 0.33	0.278					
Button-up shirt/blouse	0.04	0.16	-0.01 – 0.09	-0.05 – 0.36	0.138					
Head pulled back	0.07	0.14	-0.05 – 0.19	-0.09 – 0.37	0.23					
Sloped head posture	0.09	0.16	-0.03 – 0.21	-0.07 – 0.39	0.158					
Tattoo						0.16	0.05	-0.91 – 1.23	-0.30 – 0.41	0.766
Headphones						0.22	0.17	-0.07 – 0.50	-0.05 – 0.39	0.134
Good personal cleanliness						0.38	0.28	0.00 – 0.76	0.00 – 0.56	**0.049**
Weapon						-0.38	-0.18	-1.10 – 0.33	-0.51 – 0.15	0.288
coat rack						0.34	0.16	-0.15 – 0.83	-0.07 – 0.38	0.17
High resolution image						0.02	0.02	-0.31 – 0.35	-0.23 – 0.27	0.899
Loose-fitting top						-0.12	-0.09	-0.47 – 0.23	-0.36 – 0.18	0.493
Bag						-0.27	-0.15	-0.74 – 0.21	-0.42 – 0.12	0.267
Friendly expressions						0.07	0.04	-0.58 – 0.71	-0.37 – 0.46	0.837
Cheerful expressions						0.05	0.04	-0.40 – 0.51	-0.27 – 0.34	0.82
Interested expressions						-0.09	-0.05	-0.85 – 0.66	-0.50 – 0.39	0.807
Self-assured expressions						-0.13	-0.09	-0.57 – 0.31	-0.38 – 0.21	0.55
Indifferent expressions						-0.14	-0.08	-0.63 – 0.36	-0.39 – 0.22	0.582
Raised eyebrows						0.36	0.28	0.05 – 0.68	0.04 – 0.53	**0.026**
Smiling						0.28	0.18	-0.21 – 0.76	-0.14 – 0.49	0.257
Fast mouth movements						0.23	0.19	-0.11 – 0.56	-0.09 – 0.47	0.183
Leaning forward						0.1	0.07	-0.27 – 0.47	-0.18 – 0.32	0.585
Observations	91					64				
R2 / R2 adjusted	0.215 / 0.168					0.530 / 0.356				
F-statistic (*p*)	4.65 (< **0.001**)					3.05 (**0.001**)				

	ADEPT self-rating Conscientiousness					ADEPT observer rating Conscientiousness				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	0.25	0	0.01 – 0.49	-0.19 – 0.19	**0.045**	0.92	0	-2.57 – 4.41	-0.26 – 0.26	0.589
Big distance to camera	-0.11	-0.2	-0.21 – 0.00	-0.39 – 0.00	0.052					
Book shelve	0.05	0.14	-0.03 – 0.14	-0.08 – 0.35	0.204					
Clothing item	0.05	0.14	-0.03 – 0.12	-0.08 – 0.35	0.206	-0.12	-0.13	-0.51 – 0.26	-0.53 – 0.27	0.51
Good personal cleanliness						0.14	0.13	-0.31 – 0.58	-0.29 – 0.55	0.53
Wrinkled clothes						-0.1	-0.09	-0.48 – 0.27	-0.44 – 0.25	0.578
Much skin visible						-0.42	-0.31	-0.97 – 0.13	-0.72 – 0.10	0.13
Athletic equipment						-0.14	-0.07	-1.09 – 0.82	-0.53 – 0.40	0.767
drawer						-0.02	-0.02	-0.46 – 0.41	-0.48 – 0.43	0.913
Dyed hair						-0.22	-0.25	-0.55 – 0.10	-0.62 – 0.12	0.17
High resolution image						0.24	0.22	-0.23 – 0.71	-0.21 – 0.66	0.301
Plain room						-0.03	-0.03	-0.49 – 0.43	-0.47 – 0.41	0.88
Loose-fitting top						0.06	0.06	-0.45 – 0.57	-0.43 – 0.55	0.801
Friendly expressions						0.42	0.34	-0.65 – 1.48	-0.53 – 1.22	0.424
Cheerful expressions						0.38	0.32	-0.31 – 1.07	-0.26 – 0.89	0.264
Interested expressions						-0.46	-0.36	-1.39 – 0.47	-1.09 – 0.37	0.319
Self-assured expressions						0.18	0.14	-0.53 – 0.90	-0.40 – 0.68	0.599
Calm expressions						-0.01	0	-0.53 – 0.52	-0.44 – 0.43	0.982
Strong expressions						-0.05	-0.05	-0.65 – 0.54	-0.55 – 0.46	0.853
Tinkling with eyelashes						0.63	0.29	-0.24 – 1.49	-0.11 – 0.70	0.147
Furrowed eyebrows						-0.08	-0.07	-0.52 – 0.37	-0.49 – 0.35	0.72
Head tilt						0.15	0.14	-0.42 – 0.71	-0.39 – 0.66	0.591
Smiling						-0.43	-0.37	-1.02 – 0.16	-0.89 – 0.14	0.145
Looking at camera						0.02	0.02	-0.48 – 0.52	-0.43 – 0.48	0.926
Fast mouth movements						-0.25	-0.26	-0.67 – 0.17	-0.70 – 0.18	0.229
Relaxed posture						0.23	0.2	-0.34 – 0.81	-0.29 – 0.68	0.406
Open posture						0.6	0.54	0.01 – 1.19	0.01 – 1.07	**0.047**
Observations	97					46				
R2 / R2 adjusted	0.102 / 0.073					0.653 / 0.257				
F-statistic (*p*)	3.53 (**0.018**)					1.65 (0.125)				

	ADEPT self-rating Extraversion					ADEPT observer rating Extraversion				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	-0.33	0	-0.82 – 0.17	-0.19 – 0.19	0.192	2.34	3.45	-1.30 – 5.98	3.21 – 3.70	0.197
Book	-0.02	-0.06	-0.12 – 0.08	-0.34 – 0.22	0.654					
Map	-0.07	-0.08	-0.27 – 0.12	-0.31 – 0.14	0.458					
file cabinet	-0.04	-0.1	-0.17 – 0.09	-0.40 – 0.21	0.538					
Calm expressions	0.07	0.14	-0.04 – 0.19	-0.07 – 0.36	0.196					
Head oriented away from camera	0.11	0.2	-0.01 – 0.22	-0.02 – 0.41	0.069					
Leaning backward	0.11	0.18	-0.02 – 0.23	-0.03 – 0.39	0.087					
Painted wall	-0.06	-0.17	-0.14 – 0.01	-0.37 – 0.04	0.108					
Shutters	0.05	0.13	-0.03 – 0.14	-0.08 – 0.33	0.214					
Tattoo						-0.26	-0.11	-1.46 – 0.94	-0.60 – 0.39	0.663
Good personal cleanliness						0.22	0.21	-0.15 – 0.59	-0.14 – 0.55	0.226
Wrinkled clothes						-0.46	-0.39	-0.90 – -0.03	-0.76 – -0.02	**0.038**
Weapon						-0.05	-0.03	-0.97 – 0.88	-0.56 – 0.51	0.916
Plucked eyebrows						0	-0.01	-0.29 – 0.28	-0.41 – 0.40	0.976
Clothing item						-0.01	-0.02	-0.43 – 0.40	-0.44 – 0.41	0.942
Dark light						0.29	0.26	-0.15 – 0.74	-0.13 – 0.65	0.189
Face lit by screen						0.09	0.08	-0.29 – 0.46	-0.27 – 0.42	0.647
Heavy make up						-0.45	-0.26	-1.56 – 0.66	-0.88 – 0.37	0.411
Made-up eyes						0.31	0.27	-0.25 – 0.87	-0.22 – 0.77	0.266
Friendly expressions						0.07	0.06	-0.93 – 1.08	-0.73 – 0.84	0.879
Cheerful expressions						0.13	0.1	-0.51 – 0.76	-0.39 – 0.59	0.687
Interested expressions						-0.1	-0.08	-0.99 – 0.79	-0.76 – 0.61	0.818
Self-assured expressions						-0.06	-0.05	-0.67 – 0.54	-0.47 – 0.38	0.829
Indifferent expressions						-0.62	-0.5	-1.31 – 0.07	-1.06 – 0.06	0.077
Strong expressions						0.23	0.16	-0.37 – 0.82	-0.27 – 0.59	0.443
Tinkling with eyelashes						-0.85	-0.46	-2.58 – 0.89	-1.41 – 0.49	0.325
Furrowed eyebrows						0.31	0.25	-0.32 – 0.93	-0.26 – 0.76	0.319
Biting nails						0.49	0.11	-2.47 – 3.45	-0.53 – 0.74	0.734
Sloped head posture						0.31	0.2	-0.61 – 1.24	-0.40 – 0.80	0.495
Laughing						0.57	0.36	-0.70 – 1.83	-0.44 – 1.16	0.362
Smiling						-0.05	-0.04	-0.71 – 0.61	-0.51 – 0.44	0.88
Fast mouth movements						0.09	0.08	-0.41 – 0.60	-0.37 – 0.53	0.71
Open posture						0.06	0.05	-0.48 – 0.60	-0.40 – 0.50	0.815
Body tilted						-0.34	-0.24	-0.92 – 0.25	-0.66 – 0.18	0.245
Wrinkled forehead						0.05	0.04	-0.50 – 0.59	-0.46 – 0.54	0.863
Complete person and background visible						0.32	0.31	-0.58 – 1.22	-0.55 – 1.17	0.467
Observations	88					53				
R2 / R2 adjusted	0.249 / 0.173					0.659 / 0.290				
F-statistic (*p*)	3.28 (**0.002**)					1.79 (0.074)				

	ADEPT self-rating Agreeableness					ADEPT observer rating Agreeableness				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	-0.28	0	-0.72 – 0.17	-0.21 – 0.21	0.22	-0.89	0	-3.80 – 2.01	-0.23 – 0.23	0.538
Good personal cleanliness	0.02	0.07	-0.06 – 0.10	-0.22 – 0.35	0.647	0.44	0.38	0.03 – 0.86	0.03 – 0.73	**0.035**
Book	-0.05	-0.22	-0.10 – 0.00	-0.46 – 0.02	0.073					
coat rack	0.15	0.31	0.04 – 0.26	0.09 – 0.52	**0.007**					
Hair cover face	0.04	0.19	-0.01 – 0.08	-0.04 – 0.42	0.104					
High resolution image	0.05	0.16	-0.02 – 0.12	-0.08 – 0.40	0.189	0.02	0.02	-0.32 – 0.37	-0.26 – 0.30	0.884
Tidy room	0.05	0.23	-0.00 – 0.11	-0.01 – 0.47	0.064					
T-Shirt	-0.01	-0.07	-0.06 – 0.03	-0.33 – 0.19	0.598					
Pullover	0.02	0.1	-0.03 – 0.07	-0.14 – 0.34	0.414					
Diverse facial expressions	-0.04	-0.11	-0.21 – 0.14	-0.61 – 0.40	0.672					
Rapidly changing facial expressions	-0.01	-0.02	-0.22 – 0.20	-0.58 – 0.53	0.934					
Narrow eyes	-0.02	-0.07	-0.11 – 0.07	-0.35 – 0.22	0.644					
Painted wall	-0.02	-0.07	-0.07 – 0.04	-0.30 – 0.15	0.52					
Electronic equipment						-0.13	-0.13	-0.42 – 0.15	-0.40 – 0.15	0.356
chair						0.2	0.24	-0.01 – 0.42	-0.01 – 0.49	0.064
Kitchen						0.17	0.15	-0.12 – 0.46	-0.10 – 0.40	0.231
Loose-fitting top						0.12	0.1	-0.24 – 0.47	-0.21 – 0.42	0.515
Friendly expressions						0.11	0.08	-0.50 – 0.71	-0.36 – 0.51	0.717
Cheerful expressions						0.14	0.1	-0.41 – 0.69	-0.31 – 0.52	0.616
Interested expressions						0.09	0.06	-0.54 – 0.72	-0.36 – 0.48	0.781
Self-assured expressions						-0.12	-0.09	-0.53 – 0.28	-0.38 – 0.20	0.542
Raised eyebrows						0.33	0.27	-0.02 – 0.67	-0.02 – 0.57	0.065
Laughing						0.12	0.06	-0.57 – 0.80	-0.31 – 0.43	0.731
Smiling						0.22	0.17	-0.29 – 0.73	-0.22 – 0.56	0.383
Fast mouth movements						-0.25	-0.23	-0.64 – 0.13	-0.58 – 0.12	0.189
Open posture						0.3	0.23	-0.11 – 0.72	-0.09 – 0.55	0.15
Camera shaking						0.01	0.01	-0.37 – 0.39	-0.29 – 0.30	0.954
Observations	69					53				
R2 / R2 adjusted	0.394 / 0.264					0.548 / 0.347				
F-statistic (*p*)	3.03 (**0.002**)					2.73 (**0.006**)				

	ADEPT self-rating Neuroticism					ADEPT observer rating Neuroticism				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	0.13	0	0.02 – 0.25	-0.19 – 0.19	**0.022**	1.79	0	0.38 – 3.21	-0.18 – 0.18	**0.014**
Finished wall	0.07	0.24	0.01 – 0.13	0.04 – 0.43	**0.019**					
Painted wall	-0.06	-0.23	-0.11 – -0.01	-0.43 – -0.03	**0.022**					
Good personal cleanliness						0.11	0.1	-0.11 – 0.33	-0.09 – 0.29	0.312
Big distance to camera						0.18	0.11	-0.14 – 0.50	-0.09 – 0.31	0.27
Friendly expressions						0.24	0.2	-0.21 – 0.69	-0.17 – 0.56	0.288
Cheerful expressions						0.37	0.29	0.01 – 0.72	0.01 – 0.57	**0.042**
Interested expressions						-0.18	-0.15	-0.63 – 0.26	-0.52 – 0.22	0.417
Self-assured expressions						-0.07	-0.06	-0.39 – 0.24	-0.33 – 0.21	0.643
Calm expressions						0.32	0.25	0.06 – 0.58	0.05 – 0.45	**0.017**
Arrogant-amused expressions						-0.71	-0.34	-1.13 – -0.29	-0.54 – -0.14	**0.001**
Indifferent expressions						-0.05	-0.04	-0.34 – 0.24	-0.28 – 0.20	0.745
Smiling						0.09	0.06	-0.26 – 0.43	-0.20 – 0.33	0.628
Body tilted						0.19	0.16	-0.05 – 0.44	-0.04 – 0.37	0.124
Only Head/face visible						-0.13	-0.18	-0.26 – 0.01	-0.38 – 0.01	0.061
Observations	97					93				
R2 / R2 adjusted	0.096 / 0.077					0.371 / 0.276				
F-statistic (*p*)	4.99 (**0.009**)					3.93 (< **0.001**)				

Regression Models ADEPT (NLC)

	ADEPT self-rating Openness					ADEPT NLC rating Openness				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	-0.33	0	-0.56 – -0.11	-0.19 – 0.19	**0.004**	0.44	0	0.20 – 0.69	-0.26 – 0.26	**0.001**
Button-up shirt/blouse	0.04	0.16	-0.01 – 0.09	-0.05 – 0.36	0.138					
desk	0.06	0.12	-0.05 – 0.17	-0.10 – 0.33	0.278					
Electronic equipment	0.08	0.25	0.02 – 0.14	0.06 – 0.45	**0.013**					
Head pulled back	0.07	0.14	-0.05 – 0.19	-0.09 – 0.37	0.23					
Sloped head posture	0.09	0.16	-0.03 – 0.21	-0.07 – 0.39	0.158					
Wrinkled clothes						-0.04	-0.25	-0.08 – 0.01	-0.55 – 0.05	0.101
Asymmetrical haircut						-0.01	-0.14	-0.04 – 0.02	-0.45 – 0.17	0.359
Made-up eyes						0.02	0.12	-0.03 – 0.07	-0.16 – 0.40	0.407
Earrings						-0.02	-0.14	-0.07 – 0.03	-0.43 – 0.15	0.34
Tidy room						0.01	0.07	-0.03 – 0.04	-0.24 – 0.38	0.652
Shelve						-0.01	-0.08	-0.04 – 0.02	-0.39 – 0.23	0.616
Narrow eyes						-0.03	-0.21	-0.07 – 0.01	-0.52 – 0.10	0.171
Maintained posture						-0.05	-0.32	-0.09 – -0.00	-0.62 – -0.01	**0.04**
Open posture						0.03	0.18	-0.02 – 0.08	-0.13 – 0.48	0.245
Observations	91					49				
R^2 / R^2 adjusted	0.215 / 0.168					0.328 / 0.173				
F-statistic (*p*)	4.65 (**<.001**)					2.12 (0.052)				

	ADEPT self-rating Conscientiousness					ADEPT NLC rating Conscientiousness				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	0.25	0	0.01 – 0.49	-0.19 – 0.19	**0.045**	0.31	0	0.12 – 0.50	-0.21 – 0.21	**0.002**
Big distance to camera	-0.11	-0.2	-0.21 – 0.00	-0.39 – 0.00	0.052					
Book shelve	0.05	0.14	-0.03 – 0.14	-0.08 – 0.35	0.204					
Clothing item	0.05	0.14	-0.03 – 0.12	-0.08 – 0.35	0.206					
Headphones						-0.04	-0.18	-0.09 – 0.01	-0.39 – 0.03	0.09
Vertical picture frame						0.05	0.17	-0.01 – 0.12	-0.05 – 0.40	0.127
Arrogant-amused expressions						-0.07	-0.16	-0.16 – 0.03	-0.37 – 0.06	0.157
Adjust clothing						0.15	0.26	0.02 – 0.29	0.04 – 0.49	**0.023**
Observations	97					79				
R2 / R2 adjusted	0.102 / 0.073					0.184 / 0.140				
F-statistic (*p*)	3.53 (**0.018**)					4.17 (**0.004**)				

	ADEPT self-rating Extraversion					TSQ_Inter				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	-0.33	0	-0.82 – 0.17	-0.19 – 0.19	0.192	0.03	0	-0.43 – 0.50	-0.22 – 0.22	0.882
Shutters	0.05	0.13	-0.03 – 0.14	-0.08 – 0.33	0.214					
file cabinet	-0.04	-0.1	-0.17 – 0.09	-0.40 – 0.21	0.538					
Painted wall	-0.06	-0.17	-0.14 – 0.01	-0.37 – 0.04	0.108					
Book	-0.02	-0.06	-0.12 – 0.08	-0.34 – 0.22	0.654					
Map	-0.07	-0.08	-0.27 – 0.12	-0.31 – 0.14	0.458					
Calm expressions	0.07	0.14	-0.04 – 0.19	-0.07 – 0.36	0.196					
Leaning backward	0.11	0.18	-0.02 – 0.23	-0.03 – 0.39	0.087					
Head oriented away from camera	0.11	0.2	-0.01 – 0.22	-0.02 – 0.41	0.069					
Long hair						-0.02	-0.18	-0.04 – 0.01	-0.41 – 0.06	0.144
Glasses						0.02	0.22	-0.00 – 0.04	-0.01 – 0.46	0.063
Headset						-0.06	-0.36	-0.10 – -0.02	-0.59 – -0.12	**0.003**
Complete face visible						0.04	0.35	0.01 – 0.07	0.10 – 0.60	**0.006**
Shelve						-0.01	-0.05	-0.04 – 0.03	-0.32 – 0.21	0.689
Book shelve						-0.02	-0.14	-0.06 – 0.02	-0.42 – 0.13	0.301
Collection						-0.01	-0.04	-0.10 – 0.08	-0.32 – 0.23	0.753
Toy						-0.01	-0.03	-0.07 – 0.06	-0.32 – 0.25	0.824
Standing						0.24	0.13	-0.19 – 0.67	-0.11 – 0.36	0.274
Observations	88					61				
R2 / R2 adjusted	0.249 / 0.173					0.368 / 0.257				
F-statistic (*p*)	3.28 (**0.002**)					3.30 (**0.003**)				

	ADEPT self-rating Agreeableness					ADEPT NLC rating Agreeableness				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	-0.28	0	-0.72 – 0.17	-0.21 – 0.21	0.22	-0.14	0	-0.49 – 0.21	-0.21 – 0.21	0.418
Good personal cleanliness	0.02	0.07	-0.06 – 0.10	-0.22 – 0.35	0.647	0	-0.01	-0.05 – 0.05	-0.34 – 0.31	0.927
T-Shirt	-0.01	-0.07	-0.06 – 0.03	-0.33 – 0.19	0.598					
Pullover	0.02	0.1	-0.03 – 0.07	-0.14 – 0.34	0.414					
Hair cover face	0.04	0.19	-0.01 – 0.08	-0.04 – 0.42	0.104					
High resolution image	0.05	0.16	-0.02 – 0.12	-0.08 – 0.40	0.189	0.03	0.21	-0.01 – 0.07	-0.07 – 0.48	0.14
Tidy room	0.05	0.23	-0.00 – 0.11	-0.01 – 0.47	0.064					
coat rack	0.15	0.31	0.04 – 0.26	0.09 – 0.52	**0.007**					
Blinds	-0.02	-0.07	-0.07 – 0.04	-0.30 – 0.15	0.52					
Book	-0.05	-0.22	-0.10 – 0.00	-0.46 – 0.02	0.073					
Diverse facial expressions	-0.04	-0.11	-0.21 – 0.14	-0.61 – 0.40	0.672					
Rapidly changing facial expressions	-0.01	-0.02	-0.22 – 0.20	-0.58 – 0.53	0.934					
Narrow eyes	-0.02	-0.07	-0.11 – 0.07	-0.35 – 0.22	0.644					
Loose-fitting top						-0.02	-0.15	-0.06 – 0.02	-0.46 – 0.15	0.322
Plucked eyebrows						0.02	0.23	-0.00 – 0.05	-0.03 – 0.48	0.082
Highlights						0.04	0.11	-0.04 – 0.12	-0.13 – 0.36	0.356
Necklace						-0.03	-0.13	-0.09 – 0.03	-0.39 – 0.13	0.331
Overexposed						0.02	0.19	-0.01 – 0.04	-0.08 – 0.45	0.162
Curtains						0	0	-0.03 – 0.03	-0.26 – 0.26	0.983
bed						-0.01	-0.09	-0.05 – 0.03	-0.38 – 0.20	0.529
Stereo stand						-0.01	-0.03	-0.08 – 0.07	-0.35 – 0.29	0.86
Electronic equipment						-0.01	-0.06	-0.05 – 0.03	-0.37 – 0.25	0.69
Athletic equipment						0	0.01	-0.09 – 0.09	-0.39 – 0.40	0.964
Clothing item						-0.02	-0.16	-0.06 – 0.02	-0.46 – 0.13	0.261
Pillow						-0.02	-0.15	-0.07 – 0.03	-0.50 – 0.19	0.378
Friendly expressions						0.03	0.18	-0.02 – 0.08	-0.11 – 0.47	0.218
Self-assured expressions						0.02	0.11	-0.03 – 0.07	-0.17 – 0.39	0.442
Timid expressions						0	0.01	-0.07 – 0.07	-0.32 – 0.33	0.969
Heavy blinking						-0.02	-0.11	-0.07 – 0.03	-0.40 – 0.17	0.423
Pressed lips						-0.02	-0.14	-0.06 – 0.02	-0.40 – 0.13	0.313
Observations	69					56				
R2 / R2 adjusted	0.394 / 0.264					0.588 / 0.371				
F-statistic (*p*)	3.03 (**0.002**)					2.71 (**0.005**)				

	ADEPT self-rating Neuroticism					ADEPT NLC rating Neuroticism				
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*p*
(Intercept)	0.13	0	0.02 – 0.25	-0.19 – 0.19	**0.022**	0.34	0	-0.49 – 1.17	-0.31 – 0.31	0.4
Painting on wall	0.07	0.24	0.01 – 0.13	0.04 – 0.43	**0.019**					
Poster or picture on wall	-0.06	-0.23	-0.11 – -0.01	-0.43 – -0.03	**0.022**					
Good personal cleanliness						-0.05	-0.39	-0.15 – 0.04	-1.11 – 0.33	0.27
Athletic clothes						0	0.05	-0.07 – 0.08	-0.67 – 0.76	0.893
Loose-fitting top						-0.03	-0.22	-0.09 – 0.03	-0.69 – 0.26	0.35
Distinctive hair						-0.03	-0.23	-0.08 – 0.03	-0.69 – 0.23	0.305
Jewellery						-0.03	-0.14	-0.14 – 0.07	-0.58 – 0.30	0.515
Light halo						0.04	0.36	-0.03 – 0.12	-0.27 – 0.99	0.244
Finished wall						0.01	0.07	-0.06 – 0.07	-0.85 – 0.99	0.873
Wall pattern						-0.02	-0.22	-0.07 – 0.03	-0.70 – 0.25	0.337
Stationary						-0.05	-0.2	-0.18 – 0.08	-0.72 – 0.32	0.434
Electronic equipment						0.01	0.08	-0.06 – 0.08	-0.43 – 0.59	0.743
Weapon						-0.02	-0.09	-0.17 – 0.12	-0.67 – 0.50	0.763
Wide open eyes						0.02	0.17	-0.08 – 0.13	-0.66 – 1.01	0.67
Narrow eyes						0.01	0.05	-0.07 – 0.08	-0.53 – 0.64	0.848
Duckface						-0.12	-0.35	-0.34 – 0.10	-0.98 – 0.29	0.268
Maintained posture						0.04	0.27	-0.03 – 0.10	-0.20 – 0.73	0.247
Head pulled back						-0.03	-0.17	-0.13 – 0.07	-0.70 – 0.37	0.519
Head oriented away from camera						0.02	0.12	-0.07 – 0.11	-0.45 – 0.69	0.669
Shaking head						-0.03	-0.26	-0.11 – 0.04	-0.88 – 0.36	0.391
Shoulder movement						-0.03	-0.19	-0.09 – 0.04	-0.70 – 0.31	0.43
Touching face						-0.04	-0.35	-0.11 – 0.02	-0.90 – 0.19	0.193
Observations	97					42				
R2 / R2 adjusted	0.096 / 0.077					0.516 / 0.055				
F-statistic (*p*)	4.99 (**0.009**)					1.12 (0.399)				

XVII – Lens Models per personality trait

	cue validity		**Conscientiousness**		cue utilization	
	IPIP	ADEPT	Cue	IPIP	ADEPT obs.	ADEPT NLC
Appearance		0.20*	Clothing item		-0.28**	
			Colourful clothes	0.21*		
			Distinctive hair	-0.23*		
			Dyed hair		-0.26*	
			Good personal cleanliness	0.26**	0.44***	
			Headphones			-0.26*
	-0.33**		Long hair			
			Loose-fitting top		-0.44***	
			Much skin visible		-0.25*	
	0.23*		Nose piercing			
			Wrinkled clothes		-0.27*	
Body			Adjust clothing			0.23*
	-0.31*		Empathetic gesture clapping			
			Head tilt		0.26*	
			Open posture		0.32**	
			Relaxed posture		0.24*	
			Sitting	0.23*		
			Straight posture	0.25*		
Environment			Athletic equipment	-0.24*	-0.26**	
	0.21*		Bag			
			Bedroom	-0.39***		
		0.22*	Book shelve			
	-0.20*		chair			
			cookware & pots	0.23*		
			drawer		-0.28**	
			Pillow	-0.22*		
			Plain room		0.26*	
			Shelve	0.20*		
			Stationary	0.23*		
			Toy	0.21*		
Face			Arrogant-amused expressions			-0.20*
			Calm expressions		0.34**	
			Cheerful expressions		0.21*	
			Fast mouth movements		0.26*	
			Friendly expressions		0.39***	
			Furrowed eyebrows		0.26**	
			Interested expressions		0.37***	
	-0.22*		Laughing			
			Looking at camera	0.37***	0.26**	
			Looking down	-0.32**		
			Pausing	-0.21*		
			Self-assured expressions		0.44***	
			Smiling		0.23*	
			Strong expressions	0.22*	0.21*	
			Timid expressions	-0.25*		
			Tinkling with eyelashes		0.24*	
Media Properties			High resolution image		0.38***	
		-0.23*	Big distance to camera			
			Only parts of face	-0.41***		
			Overexposed	0.20*		
			Vertical picture frame			0.31**
	R^2	R^2		R^2	R^2	R^2
	0.18	0.10		0.61	0.59	0.22
	R^2 Adj.	R^2 Adj.		R^2 Adj.	R^2 Adj.	R^2 Adj.
	0.13	0.07		0.47	0.41	0.17

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	cue validity		**Extraversion**	cue utilization		
	IPIP	ADEPT	Cue	IPIP	ADEPT obs.	ADEPT NLC
Appearance			Clothing item		-0.24*	
			Colourful clothes	0.21*		
			Glasses			0.21*
			Good personal cleanliness		0.31**	
			Headset			-0.22*
			Heavy make up		0.31**	
	-0.30**		**Long hair**	-0.33**		-0.27*
			Made-up eyes		0.28**	
			Plucked eyebrows		0.24*	
			Pullover	-0.23*		
			Tattoo		-0.26*	
			Wrinkled clothes		-0.25*	
Body			Biting nails		0.23*	
			Body tilted		0.24*	
			Open posture		0.23*	
	-0.21*		Shaking head			
	-0.26*		Sloped head posture		0.29**	
			Standing			0.22*
Environment	-0.21*		Athletic equipment			
		-0.21*	Book			
			Book shelve			-0.22*
			Collection			-0.33**
			cookware & pots	0.24*		
			Decoration	0.21*		
		-0.23*	file cabinet			
		-0.25*	Map			
		-0.23*	Painted wall			
			Shelve			-0.21*
		0.23*	Shutters			
			Toy			-0.22*
	-0.22*		Wardrobe/closet			
			Weapon		-0.24*	
	0.23*		Window	-0.21*		
Face		0.21*	Calm expressions			
			Cheerful expressions		0.39***	
			Diverse facial expressions	0.25*		
			Fast mouth movements		0.28**	
			Friendly expressions		0.45***	
			Furrowed eyebrows		0.20*	
			Indifferent expressions		-0.27*	
			Interested expressions		0.44***	
			Laughing		0.32**	
	-0.23*		Pausing			
	-0.29**		Pout			
	-0.32**		Raised eyebrows			
			Rapidly changing facial expressions	0.25*		
	-0.22*		Rolling eyes			
			Self-assured expressions		0.38***	
	-0.21*		Sceptical expressions			
			Smiling		0.45***	
			Strong expressions	0.29**	0.32**	
	-0.26*		Swallowing hard			
			Timid expressions	-0.34**		
			Tinkling with eyelashes		0.24*	
	-0.23*		Wide open eyes			
	-0.26*		Wrinkled forehead		0.24*	
Media Properties	0.25*		Camera level with head			
			Complete face			0.30**
			Complete person and background		0.22*	
	0.20*		Dark light		0.21*	
			Face lit by screen		0.21*	
	-0.21*		Overexposed			
	R^2	R^2		R^2	R^2	R^2
	0.26	0.25		0.34	0.61	0.21
	R^2 Adj.	R^2 Adj.		R^2 Adj.	R^2 Adj.	R^2 Adj.
	0.08	0.17		0.25	0.23	0.12

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	cue validity		**Agreeableness**		cue utilization	
	IPIP	ADEPT	Cue	IPIP	ADEPT obs.	ADEPT NLC
Appearance			Business-like clothes	0.27*		
			Clothing item			-0.28*
			Distinctive clothes	-0.21*		
			Distinctive hair	-0.25*		
		0.28**	**Good personal cleanliness**		0.36***	0.32**
		0.21*	Hair cover face			
			Highlights			0.23*
			Loose-fitting top	-0.30*	-0.25*	-0.36**
	0.20*		Necklace			-0.21*
			Plucked eyebrows			0.23*
		0.22*	Pullover			
		-0.20*	T-Shirt			
Body			Arms behind back	-0.31*		
			Body tilted	-0.27**		
			Body turned away	-0.27**		
			Empathetic gesture clapping	-0.30*		
			Head oriented away from camera	-0.28**		
			Leaning forward	-0.28**		
			Shaking legs			-0.84*
			Shoulder movement	-0.28*		
			Sloped head posture	-0.26**		
			Slouched posture	-0.24*		
			Straight posture	0.29**		
			Touching body	-0.29**		
			Touching hair	-0.22*		
			Touching head	-0.23*		
Environment			Athletic equipment	-0.22*		
			bed			-0.20*
		-0.27**	Book			
			chair		0.20*	
		0.22*	coat rack			
			Curtains			-0.22*
			Electronic equipment		-0.20*	-0.20*
			Kitchen		0.24*	
			Another person present	-0.24*		
		-0.20*	Painted wall			
			Pillow	-0.25*		-0.21*
	-0.25*		Shelve			
			Stereo stand	-0.20*		-0.21*
		0.29*	**Tidy room**	0.25*		

Table continues on next page

Table starts at previous page

Face			Biting lips	-0.24*		
			Cheerful expressions		0.32**	
		-0.22*	Diverse facial expressions			
			Duckface	-0.21*		
			Friendly expressions		0.46***	0.23*
			Furrowed eyebrows	-0.29**		
			Heavy blinking			-0.23*
			Interested expressions		0.39***	
	-0.23*		Licking lips			
		-0.27**	Narrow eyes			
			Pausing	-0.32**		
	-0.22*		**Plopping lips**	-0.21*		
			Pout	-0.21*		
			Pressed lips	-0.34**		-0.28*
		-0.23*	Rapidly changing facial expres-sions			
			Rolling eyes	-0.20*		
			Self-assured expressions		0.21*	0.27**
	-0.21*		Surprised expressions			
			Timid expressions	-0.30**		-0.29**
	-0.22*		Wide open mouth			
Media Properties		0.24*	**High resolution image**		0.31**	0.22*
			Big distance to wall	-0.20*		
	0.23*		Dark light			
	0.25*		Even light			
			Only parts of face	-0.25*		
			Overexposed	0.30**		0.26*
	R^2	R^2		R^2	R^2	R^2
	0.32	0.36		0.62	0.47	0.55
	R^2 Adj.	R^2 Adj.		R^2 Adj.	R^2 Adj.	R^2 Adj.
	0.24	0.27		0.35	0.34	0.38

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

	cue validity		**Neuroticism**	cue utilization		
	IPIP	ADEPT	Cue	IPIP	ADEPT obs.	ADEPT NLC
Appearance			Asymmetrical haircut	-0.25*		
			Athletic clothes			-0.22*
			Distinctive hair			-0.24*
	0.27*		Dyed hair			
			Good personal cleanliness	0.25*	0.23*	0.21*
			Hair cover face	-0.30**		
			Heavy make up	0.22*		
			Jewellery			-0.21*
			Tattoo			-0.30**
			Tended hair	0.24*		
Body			Body tilted		0.23*	
			Head oriented away from camera			-0.26*
			Head pulled back			-0.24*
			Head tilt	0.21*		
			Maintained posture			0.32**
	-0.21*		**Shaking head**			-0.24*
			Shoulder movement			-0.34**
			Slouched posture	0.25*		
			Touching face			-0.22*
Environment			Electronic equipment			-0.24*
		0.209*	**Finished wall**			0.25*
			Kitchen	0.26*		
		-0.203*	Painted wall			
			Shelve	-0.28**		
			Stationary			-0.21*
			Stereo stand			-0.24*
			Wall pattern			-0.23*
			Weapon			-0.39***
Face			Arrogant-amused expressions		-0.20*	
	-0.22*		Calm expressions		0.20*	
			Cheerful expressions		0.28**	
			Duckface			-0.33**
			Friendly expressions		0.38***	
			Indifferent expressions		-0.20*	
			Interested expressions		0.32**	
	-0.29**		Looking at camera			
			Narrow eyes			-0.29**
	-0.26*		Raised eyebrows			
			Self-assured expressions		0.23*	
			Smiling		0.29**	
			Surprised expressions	-0.21*		
	-0.25*		Timid expressions			
			Wide open eyes			-0.20*
	-0.23*		Wrinkled forehead			
Media Properties			Light from right side	-0.22*		
			Big distance to camera		0.21*	
	0.30**		Dark light			
			Even light	0.20*		
	-0.22*		Face lit by screen			
			Light halo			-0.23*
			Only Head/face visible		-0.22*	
	-0.20*		Vertical picture frame			
	R^2	R^2		R^2	R^2	R^2
	0.4	0.1		0.41	0.4	0.37
	R^2 Adj.	R^2 Adj.		R^2 Adj.	R^2 Adj.	R^2 Adj.
	0.3	0.08		0.32	0.29	0.08

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

XVIII – Factor analyses

	F1	F2	F3	F4	F5	F6	Communality	Uniqueness	Complexity
Body turned away	0.79	*0.24*	*0.13*	*-0.03*	*0.21*	*0.35*	0.87	0.13	1.83
Tinkling with eyelashes	0.78	*0.01*	*0.29*	*0.14*	*0.13*	*0.23*	0.79	0.21	1.61
Laughing	0.71	*0.10*	*0.15*	*0.41*	*0.16*	*-0.05*	0.72	0.28	1.88
Swiveling on chair	0.71	*0.26*	*0.06*	*-0.07*	*0.22*	*0.24*	0.68	0.32	1.78
Biting lips	0.69	*0.20*	*0.36*	*0.05*	*0.16*	*0.11*	0.68	0.32	1.92
Touching head	0.65	*0.12*	*0.10*	*0.12*	*0.07*	*0.45*	0.67	0.33	2.01
Touching hair	0.62	*0.25*	*0.08*	*-0.08*	*0.17*	*0.34*	0.6	0.4	2.18
Head pulled back	0.59	*0.33*	*0.32*	*0.07*	*-0.03*	*0.06*	0.58	0.42	2.28
Heavy blinking	0.52	*0.50*	*0.17*	*0.02*	*0.05*	*-0.02*	0.55	0.45	2.24
Grumpy expressions	0.51	*0.24*	*0.22*	*-0.24*	*0.15*	*0.27*	0.51	0.49	3.25
Head oriented away from camera	0.49	*0.42*	*0.24*	*-0.02*	*-0.04*	*0.44*	0.67	0.33	3.46
Licking lips	0.48	*0.21*	*0.37*	*0.01*	*0.12*	*-0.10*	0.43	0.57	2.55
Pout	0.47	*0.45*	*0.21*	*-0.15*	*0.24*	*0.31*	0.65	0.35	3.94
Leaning backward	0.46	*0.23*	*0.20*	*-0.20*	*-0.10*	*-0.02*	0.36	0.64	2.47
Duckface	0.46	*0.28*	*0.09*	*-0.04*	*0.07*	*0.28*	0.39	0.61	2.58
Change of position	0.45	*-0.09*	*0.00*	*0.13*	*-0.05*	*0.06*	0.23	0.77	1.31
Maintained posture	-0.35	*-0.20*	*0.01*	*-0.01*	*-0.06*	*0.11*	0.18	0.82	1.92
Nodding	*0.23*	0.83	*0.00*	*0.13*	*0.20*	*-0.02*	0.8	0.2	1.33
Shaking head	*0.08*	0.78	*0.02*	*0.11*	*0.18*	*0.11*	0.66	0.34	1.22
Wide open mouth	*0.13*	0.76	*0.23*	*0.23*	*0.03*	*-0.04*	0.7	0.3	1.46
Rolling eyes	*0.26*	0.75	*0.27*	*0.03*	*0.32*	*0.03*	0.8	0.2	1.93
Wide open eyes	*0.21*	0.62	*0.26*	*0.15*	*0.03*	*0.10*	0.53	0.47	1.82
Swallowing hard	*0.25*	0.53	*0.43*	*0.05*	*0.13*	*-0.02*	0.54	0.46	2.55
Arrogant-amused expressions	*0.26*	0.52	*0.21*	*-0.03*	*0.22*	*0.28*	0.51	0.49	2.97
Looking up	*0.05*	0.51	*0.21*	*0.11*	*-0.02*	*0.28*	0.39	0.61	2.06
Fast mouth movements	*-0.06*	0.48	*0.00*	*0.28*	*-0.05*	*0.39*	0.47	0.53	2.64
Pausing	*0.16*	0.47	*0.46*	*-0.13*	*0.04*	*0.07*	0.49	0.51	2.45
Narrow eyes	*0.17*	0.47	*0.35*	*0.01*	*0.20*	*0.11*	0.43	0.57	2.72

Table continues on next page

Table starts at previous page

Closed eyes	*0.33*	0.46	*0.46*	*0.11*	*0.10*	*0.06*	0.55	0.45	3.05
Looking away from camera	*0.07*	0.39	*0.31*	*-0.06*	*-0.05*	*0.32*	0.36	0.64	3.04
Raised eyebrows	*-0.03*	0.38	*0.36*	*0.18*	*0.34*	*0.13*	0.44	0.56	3.68
Camera shaking	*0.09*	*0.29*	*0.10*	*-0.11*	*0.05*	*0.06*	0.12	0.88	1.95
Pressed lips	*0.18*	*0.30*	0.68	*0.03*	*-0.05*	*-0.03*	0.59	0.41	1.57
Furrowed eyebrows	*0.14*	*0.12*	0.62	*0.03*	*0.07*	*0.26*	0.5	0.5	1.57
Wrinkled forehead	*0.14*	*0.17*	0.60	*-0.04*	*0.21*	*0.16*	0.48	0.52	1.73
Plopping lips	*0.31*	*0.46*	0.53	*0.10*	*0.27*	*-0.10*	0.69	0.31	3.35
Sloped head posture	*0.48*	*0.19*	0.53	*0.08*	*0.16*	*0.21*	0.62	0.38	2.87
Looking down	*0.12*	*0.06*	0.47	*-0.13*	*0.00*	*0.11*	0.27	0.73	1.44
Skeptical expressions	*0.01*	*0.30*	0.47	*-0.10*	*0.27*	*0.17*	0.42	0.58	2.84
Friendly expressions	*0.19*	*0.11*	*-0.02*	0.88	*0.04*	*0.11*	0.83	0.17	1.17
Interested expressions	*-0.08*	*0.12*	*-0.02*	0.83	*0.04*	*0.34*	0.82	0.18	1.41
Cheerful expressions	*0.27*	*0.20*	*-0.01*	0.65	*0.35*	*0.02*	0.66	0.34	2.15
Smiling	*0.43*	*-0.04*	*0.08*	0.64	*-0.03*	*-0.02*	0.6	0.4	1.8
Self-assured expressions	*0.01*	*0.12*	*0.03*	0.59	*0.02*	*0.40*	0.53	0.47	1.87
Indifferent expressions	*0.44*	*-0.03*	*0.01*	-0.52	*0.04*	*-0.03*	0.47	0.53	1.97
Looking at camera	*-0.14*	*-0.02*	*-0.17*	0.46	*-0.15*	*-0.14*	0.3	0.7	1.91
Diverse facial expressions	*0.17*	*0.34*	*0.20*	*0.16*	0.78	*0.07*	0.82	0.18	1.76
Changing facial expressions	*0.06*	*0.40*	*0.27*	*0.12*	0.77	*0.13*	0.85	0.15	1.94
Straight head posture	*0.00*	*0.08*	*-0.01*	*0.13*	-0.48	*-0.06*	0.25	0.75	1.24
Surprised expressions	*0.39*	*0.23*	*0.19*	*0.04*	0.47	*0.11*	0.48	0.52	2.97
Strong expressions	*0.20*	*0.25*	*0.18*	*0.31*	0.46	*0.10*	0.45	0.55	3.35
Calm expressions	*-0.13*	*-0.05*	*0.10*	*0.10*	-0.42	*0.31*	0.31	0.69	2.38
Head tilt	*0.20*	*0.06*	*0.10*	*0.14*	*-0.02*	0.57	0.4	0.6	1.46
Body tilted	*0.28*	*0.01*	*0.10*	*0.21*	*0.28*	0.50	0.46	0.54	2.79
Leaning forward	*0.30*	*0.07*	*0.22*	*0.17*	*0.11*	0.44	0.38	0.62	2.86
Touching face	*0.16*	*0.24*	*0.17*	*0.07*	*0.36*	0.44	0.44	0.56	3.27
Timid expressions	*0.15*	*0.22*	*0.14*	*-0.06*	*0.04*	*0.25*	0.16	0.84	3.51

Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy indicates the sample including the static visual cues is unsuitable for factor analysis. KMO = 0.1523708.

Property	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
SS loadings	7.762	7.225	4.543	4.069	3.347	3.195
Proportion Var	0.136	0.127	0.080	0.071	0.059	0.056
Cumulative Var	0.136	0.263	0.343	0.414	0.473	0.529
Proportion Explained	0.258	0.240	0.151	0.135	0.111	0.106
Cumulative Proportion	0.258	0.497	0.648	0.783	0.894	1.000
Factor 1	1.000	0.018	0.023	-0.023	0.025	0.036
Factor 2	0.018	1.000	0.026	-0.018	0.028	-0.008
Factor 3	0.023	0.026	1.000	0.003	0.020	0.020
Factor 4	-0.023	-0.018	0.003	1.000	-0.006	0.051
Factor 5	0.025	0.028	0.020	-0.006	1.000	-0.006
Factor 6	0.036	-0.008	0.020	0.051	-0.006	1.000

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit allein verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe. Diese Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Großensee, 12. Oktober, 2022

Richard Justenhoven