

You Can Record an Asynchronous Video Interview Anywhere! ... But Should You?

by

Christina Scott

A Thesis Submitted to
Saint Mary's University, Halifax, Nova Scotia
In Partial Fulfillment of the Requirements for the Degree of Master of Science in Applied
Psychology
(Industrial-Organizational Psychology)

May 2022, Halifax, Nova Scotia

© Christina Scott, 2022

Approved: Dr. Nicolas Roulin
Supervisor

Approved: Dr. Kevin Kelloway
Committee member

Approved: Dr. Debra Gilin
Committee member

Approved: Dr. Johannes Basch
External

Date: May 18, 2022

Abstract

You Can Record an Asynchronous Video Interview Anywhere! ... But Should You?

By Christina Scott

Abstract: Very little is known in terms of the possible biasing effects that AVI background visual cues may introduce to the selection process. This study examined the possible biasing effects of background setting on initial impression and final interview outcomes. Using a 3 (professional versus blurred versus personal background) x2 (high versus low quality applicant responses) between-subjects design, 400 Prolific participants rated one of six manipulated AVI recorded conditions. The results showed that initial impressions influenced final interview scores, and that applicant response quality had a significant main effect on final interview scores. Background setting was not associated with initial impression or final interview scores. These findings contribute to the literature by testing some of the theoretical elements included in the dual-process theory framework. The results also complement existing practical guidance provided to organizations and applicants seeking to reduce the potential biasing effects introduced by visible background features in selection processes.

May 18, 2022

Acknowledgements

I would first like to thank my supervisor, Dr. Roulin, for all of his help and support in developing this thesis over the past year. I appreciate all of the time, energy and patience that you dedicated to helping me along this process. Without your voice of reason and responsiveness, especially when I encountered what I thought were crises, I would never have reached this stage. Thank you for allowing me to learn so much, and for helping to make this process as smooth as possible!

I would also like to extend my thanks to my committee members, Dr. Kelloway and Dr. Gilin, for their time and valuable feedback. Your insightful suggestions strengthened the theoretical and statistical foundation of my thesis, and encouraged me to critically consider interrelated factors that I otherwise would not have considered.

In addition, a special thanks goes to my classmate and colleague, Keith Joy, for recording countless mock AVI videos on my behalf. I appreciate the time and attention that you took in making sure that the videos were of a high quality, especially since you had nothing to personally gain in doing so.

Finally, I would like to thank my husband, Matt, for his incredible love and support over the past two years. Thank you for taking over the lion's share of our family and household responsibilities, and for being an amazing father to our son when I needed time to focus on academics. Without you, I never would have gotten this far, and I am looking forward to our next adventures!

Table of Contents

LITERATURE REVIEW	6
THE ADVANTAGES ASSOCIATED WITH AVIS	7
THE POTENTIAL NEGATIVE EFFECTS OF AVIS.....	9
WHY THE AVI BACKGROUND SETTING MAY BE STIGMATIZING.....	11
MECHANISMS OF BIASED JUDGEMENTS PROPOSED BY THE DUAL PROCESS THEORY	14
TYPE 1 AND TYPE 2 PROCESSES	16
INITIAL IMPRESSION FORMATION.....	21
INDIRECT EFFECTS OF INITIAL IMPRESSIONS ON INTERVIEW PERFORMANCE	23
FEELING OF RIGHTNESS	25
MODERATING EFFECTS	26
<i>Raters' Need for Cognitive Closure</i>	27
<i>Quality of Candidate Responses</i>	28
METHODS	29
DESIGN	29
PILOT STUDIES.....	31
PROCEDURE	39
PARTICIPANTS.....	40
MEASURES.....	42
RESULTS	44
DIRECT & INDIRECT EFFECTS.....	52
MODERATORS	54
FEELING OF RIGHTNESS	54

RECORD AN ASYNCHRONOUS INTERVIEW ANYWHERE	5
EXPLORATORY ANALYSES	56
DISCUSSION	60
THEORETICAL IMPLICATIONS	60
PRACTICAL IMPLICATIONS	72
LIMITATIONS	76
CONCLUSION	77
REFERENCES	80
APPENDIX A: JOB DESCRIPTION/POSTING	91
APPENDIX B: CANDIDATE RESPONSE SCRIPTS	92
APPENDIX C: BACKGROUND MANIPULATIONS	97

Figures

FIGURE 1: CONCEPTUAL MODEL OF THE STUDY DESIGN	16
FIGURE 2: CONCEPTUAL MODEL WITH UNSTANDARDIZED REGRESSION WEIGHTS AND SEs	55

Tables

TABLE 1: RESPONSE SCRIPT AND STILL PHOTO RATINGS FOR PILOT STUDIES 1A AND 1B	34
TABLE 2: BOOTSTRAPPED DESCRIPTIVE STATISTICS AND INTERCORRELATIONS	50
TABLE 3: AVERAGE INITIAL IMPRESSION, RESPONSE-LEVEL, FINAL INTERVIEW, FEELING OF RIGHTNESS, AND NFCC RATINGS ACROSS CONDITIONS.	51

Asynchronous video interviews (AVIs) have become increasingly common across selection processes worldwide (HireVue, 2019; Modern Hire, 2020), yet very little is known in terms of the potential adverse effects experienced by some applicants as a result of using this modality (Basch & Melchers, 2019; Lukacik et al., 2022). In particular, some researchers have raised concerns that the candidate backgrounds captured in AVI recordings could stigmatize certain applicants and disproportionately bias final interview performance ratings (Lukacik et al., 2022). The introduction of such bias could have debilitating implications for both organizations and applicants, as organizations may inadvertently screen out suitable candidates from consideration, and limit the diversity of their hires (Catano et al., 2022). These potential negative consequences highlight the need to better understand how background setting may affect interview performance outcomes. This study addressed this need using components of a dual-process theory framework (Derosus et al., 2016).

This study contributes to the literature, as well as informs future applicants and organizations seeking to leverage interview best-practices in AVI contexts. This study is among the first to examine background cues in AVIs by testing some of the elements included in the dual-process theory framework outlined by Derosus et al. (2016) in the context of background cues in AVIs. This study also takes some initial steps to more closely examine the mechanisms involved in the formation of biased judgements resulting from visual cues present in a technologically-mediated environment. The results of this study may help inform candidates concerned about their AVI background selection, and fall in line with existing organizational guidance in terms of how to minimize biased initial impressions on the part of interview raters and promote fair AVI selection processes. Overall, this study contributes towards the growing

literature on AVIs, and points to the need for additional research in order to better understand, and ultimately alleviate, potential sources of AVI bias.

The Advantages Associated with AVIs

AVIs represent an alternative interview modality to traditional face-to-face interviews. In contrast to face-to-face interviews, AVIs are completed entirely online and thus do not involve interpersonal interactions between the interviewer and the applicant. Instead, applicants record video responses to questions provided by the employer, which are subsequently uploaded and eventually scored by one or several interviewers, or in some cases, a computer algorithm (Hemamou et al., 2019; Lukacik et al., 2022; Naim et al., 2018).

The lack of interaction inherent to this interview modality offers a number of advantages to organizations. From an economic standpoint, AVIs are cheaper than traditional interviews, as AVIs require less employee time to conduct and there is no need to cover travel expenses to the interview site (Basch & Melchers, 2019; Lukacik et al., 2022; Mejia & Torres, 2018). In addition, AVIs require less administrative coordination, as they can be completed at the applicants' convenience, and technically from any location with an appropriate internet connection (Gorman et al., 2018; Lukacik et al., 2022). Many AVI platforms therefore advertise the fact that this interview modality allows organizations to reach a larger and more diverse selection pool, as organizations are not limited to interviewing candidates only within a given geographical region (Catano et al., 2022; Gorman et al., 2018; HireVue, 2019).

Due to their highly structured nature (i.e. Lukacik et al., 2022), AVIs have also been touted for their apparent reliability and validity-related advantages. Structured interviews are designed to promote consistency within and across interviews, which reduces rater bias, and also allows raters' to make more objective decisions in terms of which candidates are the most

suitable for a given position (Huffcutt et al., 2001; Levashina et al., 2014). For instance, some structured elements, such as reducing interviewer prompting, limiting ancillary information, and asking the same questions across applicants (Campion et al., 1997) are particularly practicable in the AVI modality (Lukacik et al., 2022; Torres & Gregory, 2018). AVIs typically do not include a rapport-building stage due to the lack of interaction between the interviewer and applicant, thus reducing the amount of biasing ancillary information being shared (see rapport building in Levashina et al., 2014). In addition, since questions are pre-recorded or scripted, AVIs also ensure that all candidates are asked identical questions. Prompting and probing on the interviewer's part is also eliminated, thus ensuring consistency in questioning across interviews (Basch & Melchers, 2019; Lukacik et al., 2022).

Based on these advantages, it is not surprising that AVIs are increasing in prevalence across selection systems worldwide. The fact that the Modern Hire AVI platform conducted over 20 million interviews and assessments and experienced a 40% increase in users in 2019 attests to the rising popularity of AVIs in modern selection processes (Modern Hire, 2020). AVIs are also being widely utilized across a variety of industries, with as many as one third of Fortune 500 companies purported to use the HireVue AVI platform alone (HireVue, 2019). It is therefore unlikely that AVIs represent a short-term passing trend. Yet, despite their widespread use, research has lagged behind in understanding the potential ramifications of conducting interviews using this modality (Basch & Melchers, 2019; Lukacik et al., 2022). As such, the question as to whether the advantages associated with AVI usage outweigh their potential adverse effects continues to be a matter of speculation.

The Potential Negative Effects of AVIs

Much of the AVI research to date has identified several disadvantages associated with AVI usage, however, little empirical work has been done to understand the potential biases that this modality may introduce. Recent meta-analytic results indicated that synchronous video-mediated interviews tend to result in lower performance scores compared to face-to-face interviews (Blacksmith et al., 2016), which suggests that certain elements inherent to technologically-mediated interviews may negatively influence interview outcomes for applicants (Lukacik et al., 2022). For example, some research has attributed the difference in performance scores to the decreased ability to view non-verbal behaviours and the limited opportunity to use impression management techniques in video modalities (Blacksmith et al., 2016). Despite these findings, however, little research to date has explored the potential adverse effects related to features of the interview modality itself. As such, several studies and research agendas have called for closer examination as to how technological features may influence performance ratings and the fairness of selection processes (Blacksmith et al., 2016; Landers & Marin, 2021; Lukacik et al., 2022). In particular, some researchers have raised concerns regarding the ability to view applicants' surrounding environment in the background of AVI recordings, which may introduce a unique dimension of bias to employment interviews (Lukacik et al., 2022).

Unlike traditional face-to-face interviews, in which an interviewer is only privy to visual information gleaned from candidates' appearance, clothing and grooming, AVI recordings offer additional visual cues captured within an applicant's background. Visible features have been associated with biased judgements in both face-to-face and AVI contexts (Suen et al., 2019). Several studies, for example, found that physically attractive AVI candidates were rated more highly or were more likely to be invited to subsequent interview stages compared to their less

attractive counterparts (Suen et al., 2019; Torres & Gregory, 2018). Because environmental features provide additional visual cues to influence initial impressions (Gosling et al., 2002), there is concern that raters may draw inferences pertaining to an individual's suitability from clues garnered from applicants' recorded surroundings (Lukacik et al., 2022). For example, AVI raters may notice personal photos, décor, furniture or other objects in the applicant's background that may signal personal information, legally protected statuses or an applicant's socioeconomic status. These additional visual cues may cause certain applicants to become inadvertently stigmatized by those evaluating their interview performance.

Given that stigmatization has been tied to lower interview performance ratings (Buijsrogge et al., 2021; Hebl & Kleck, 2002), biased perceptions could ultimately reduce some applicants' chances of selection, despite the suitability of their job-related skills and qualifications. Such adverse effects not only negatively affect applicants, but also organizations, which may inadvertently limit the diversity of their hires (Catano et al., 2022), as those falling into certain stigma groups (i.e. a protected group; Torres & Gregory, 2018) may be eliminated from consideration by virtue of their background cues. Similarly, organizations may miss out on high quality applicants, who may be well-suited for the position, but who failed to rank as highly as less-suitable candidates due to biased judgements (Catano et al., 2022). Given the potential risks and consequences associated with visual cues in AVIs, it is therefore critical to better understand the extent to which certain factors may unfairly disadvantage some applicants. To do so, it is important to not only identify which features inherent to AVIs may contribute to unfair selection decisions, but to also examine why certain elements may be stigmatizing in the first place.

Why AVI Background Setting may be Stigmatizing

Social psychology offers a fitting theoretical explanation in terms of why certain factors may be stigmatizing in an AVI context. Social psychology suggests that an individual's personal characteristics become stigmatizing when they fall outside of normative expectations and are viewed unfavourably as a flaw by others (Goffman, 2009). Individuals observing these differentiating features attribute specific stereotypical characteristics to the stigmatized person (Jones, 1984, as cited in Summers et al., 2018), resulting in the stigmatized individual being excluded or negatively impacted in some way (Summers et al., 2018). For example, in work (and arguably in AVI) contexts, such deviations from normative expectations may result in individuals being passed over for job or career advancement opportunities. It could therefore be surmised that individuals who breach normative AVI behaviours could be subject to stigmatization, and thus reduce their chances of being hired. Although some norms associated with face-to-face modalities cannot translate into an AVI context due to the lack of interaction between the interviewer and the candidate (i.e. it is impossible to shake hands), there are nevertheless some normative behaviours expected of those interacting through video modalities.

For instance, a cursory search of the internet yields countless recommendations for video conferencing etiquette, and a common theme that consistently emerges is that individuals should avoid engaging in interviews or meetings from within their bedrooms, particularly if the bed is visible to the observer (Cummins, 2021; Myrick, 2021; Vassel, 2020). The fact that similar recommendations are widespread across a variety of websites suggests that there is a societal expectation that individuals will participate in AVIs or video conferencing in a professional-looking environment, rather than a personal space. Deviations from this societal norm (i.e. recording an AVI in a bedroom) could therefore be viewed as unfavourable, thus potentially

stigmatizing the individual during the interview evaluation process. This perspective is also supported by research findings, which indicated that recruiters perceived job applicants to be unprofessional when they failed to reflect formal aesthetic qualities (Barrick et al., 2009; Torres & Gregory, 2018) or interviewed “in front of inappropriate backgrounds, photos or posters” (Mejia & Torres, 2018, p. 694).

The extent to which a stigmatizing feature or deviation from a societal norm is perceived as negative depends on a number of factors. Although some debate exists as to which dimensions of stigma most significantly influence observer’s perceptions (Heatherton et al., 2003), the distinctions of visibility and controllability (Crocker et al., 1998; Summers et al., 2018) appear most applicable to stigma resulting from background cues in AVIs. Visibility and controllability of stigma influence the extent to which individuals may be stigmatized by visual cues presented in their AVI backgrounds.

Visibility of stigma is considered to be one of the most important dimensions of stigma, as features that are readily apparent to the perceiver often serve as a schema in which other inferences are made (Crocker et al., 1998; Summers et al., 2018). Essentially, for a feature to become stigmatizing, it must be apparent to others (Goffman, 2009; Summers et al., 2018). In interview contexts, visual stigma can take many forms, although until now, research has primarily focused on physical characteristics that distinguish an individual from others (Barrick et al., 2009; Mejia & Torres, 2018). For example, research into obesity and visible disabilities (Hebl & Kleck, 2002), facial stigma (i.e. port-wine stains; Buijsrogge et al., 2021) and attractiveness (Dipboye et al., 1975; Suen et al., 2019) have repeatedly found that those with visible differences tend to obtain lower scores during interviews compared to their non-stigmatized counterparts. Although visual background cues cannot be considered a physical

characteristic of the applicants themselves, by virtue of being viewed simultaneously alongside the individual in AVI recordings, background details nevertheless present visual information that is likely attributable to the applicants themselves. For instance, a personal background may signal that an individual is unskilled, or may not be conscientious, while a professional background may point to opposite conclusions.

Similarly, the extent to which the background content is visible (i.e. blurred versus completely visible) would likely influence the degree of stigma associated with it. Goffman (2009) and Summers et al. (2018) suggest that the visibility of stigma not only refers to whether the stigma is visible to the observer, but also encompasses the extent to which the stigmatizing features are discernable. In an AVI context, the background blur functions would likely limit the visibility of any stigmatizing features, which would in turn limit the extent to which an applicant is negatively associated with the stigma. As a result, blurring stigmatizing features in the background could potentially neutralize the negative effect of stigma compared to when a background is completely visible in an AVI context.

Another important dimension of stigma is the degree of controllability that the stigmatized individual has over the stigmatizing feature, in that those who display a stigmatizing feature that is within their control (i.e. obesity) tend to be judged more negatively than those with no control over the stigmatizing feature (i.e. disability; Hebl & Kleck, 2002). For example, research found that those who dressed more casually or groomed themselves to a lower standard in job interviews were judged more negatively compared to those who wore professional attire and were well-groomed (Barrick et al., 2009; Torres & Gregory, 2018). Since the AVI background setting is at the discretion of the applicant, the visual elements presented in the background would likely be perceived as a choice in terms of how the applicants decided to

present themselves. As such, backgrounds that project a more professional image (i.e. a private office space) would likely be viewed as less stigmatizing than those displaying a less professional image (i.e. a bedroom or the outdoors). Similarly, choosing a visible stigmatizing background rather than blurring it could result in increased applicant stigma. As such, the high degree of visibility and controllability associated with certain AVI background features likely promote candidate stigmatization. This stigmatization may then lead to subsequent biased interview outcomes through mechanisms described by the dual-process theory framework.

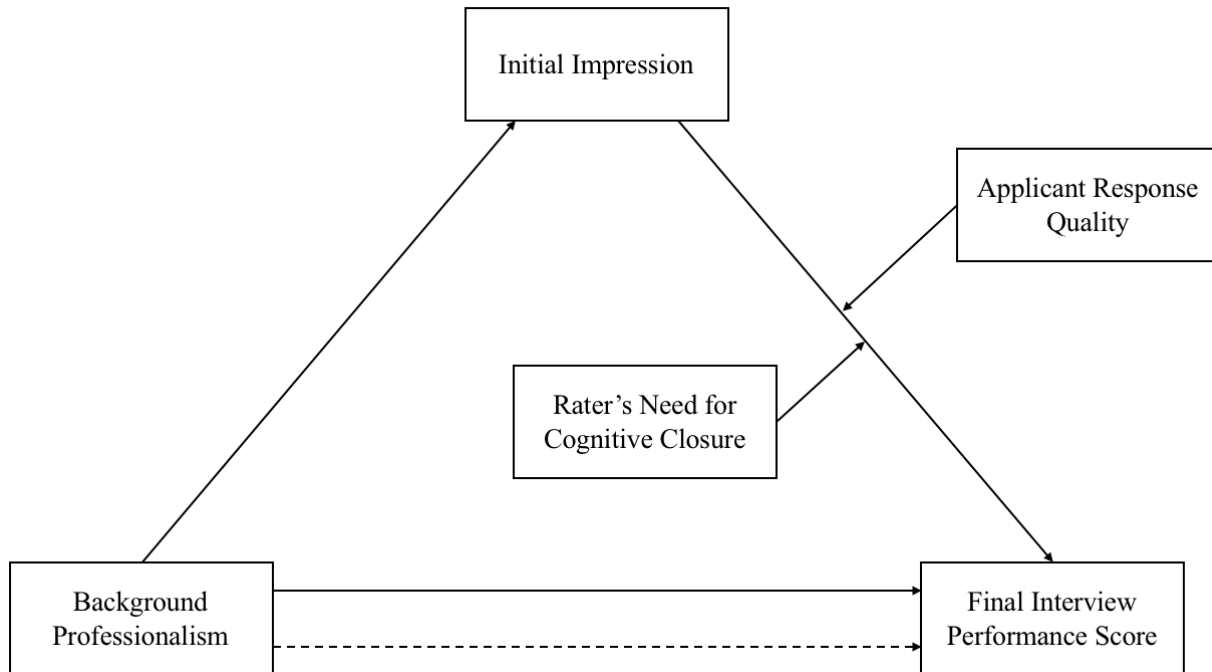
Mechanisms of Biased Judgements Proposed by the Dual Process Theory

The current study pulls elements from the interview framework introduced by Derous et al. (2016), which are depicted in Figure 1. In their model, Derous et al. (2016) proposed that stigmatized applicants' interview scores may be adversely affected through biasing mechanisms inherent to the dual-process theory (Evans, 2008). Given the multitude of dual-processing theories that exist across a number of domains (i.e. the Elaboration Likelihood Model in the social psychology field; Kitchen et al., 2014) and variations in terms of how various elements and processes are defined (i.e. Type 1 processes, heuristics, etc; Evans, 2008), theorists to date have struggled to combine the various perspectives into a single, overarching theory. As such, it is important to note that Derous et al. (2016) based their model primarily on the cognitive reasoning approach to the dual-processing theory, which specifically considers how biases are developed during decision-making or judgement formation processes (Evans, 2008).

The cognitive reasoning approach emphasizes how information is ultimately deemed relevant or irrelevant in a decision-making process, and describes how it contextualizes individual judgements, which is particularly fitting to interview contexts. Although similar in their acknowledgement of the existence of two distinct processes, other dual-processing theory

approaches are less directly applicable to the examination of bias in interview settings. For instance, dual-process theories in the social cognition field (such as the Elaboration Likelihood Model) tend to rely more on rule-based reasoning rather than unconscious judgements (Evans, 2008), and unconscious judgements are considered to be a pertinent factor in bias development (Evans, 2008; Pryor et al., 2004). As such, Derous et al.'s (2016) emphasis on the cognitive reasoning approach appears to be a logical choice for their theoretical framework focused on biased judgements in interview settings.

Derous et al.'s (2016) framework addresses a number of existing shortcomings observed in the field of interview decision-making research, and therefore advances the ways in which interview processes are understood and examined in research. For instance, rather than focusing on bias as merely an interview outcome (as was previously done), Derous et al.'s dual-process theory framework proposes that stigmatizing features might affect raters' decision-making processes throughout the course of an interview. Their framework also goes beyond depicting interview decision-making as a strictly rational decision-making process, by incorporating "interviewers' fallacies and irrationalities" (Derous et al., 2016, p. 104), which likely play an important part in final interview outcomes. Given that the framework aims to describe the development and effects of visible and non-concealable stigma in interview settings, it is therefore especially well-suited to be used as a theoretical basis for understanding the ramifications of background cues in an AVI context (Lukacik et al., 2022).

Figure 1*Conceptual Model of the Present Study Design*

Note. This model depicts the anticipated direct and indirect effects, as well as the moderating variables used in the study. The relationship between initial impression and final interview performance score is moderated by applicant response quality and rater need for cognitive closure (NFCC).

Type 1 and Type 2 Processes

In order to understand how biased decisions are made and how they influence interview outcomes, it is first necessary to differentiate between the basic processes involved in judgement formation. The dual-process theory supposes that judgements pertaining to an applicant's job suitability or interview performance are guided by two separate, but simultaneous, processes (Derous et al., 2016; Evans, 2008). Often referred to as Type 1 and Type 2 processes, these

processes are differentiated by the extent to which they rely on working memory and conscious processes (Derosus et al., 2016).

Type 1 processes represent the immediate responses that individuals experience when observing stigma (Derosus et al., 2016). Initial responses to stigma are thought to be either evolutionary in nature or the product of conditioned emotional reactions (Pryor et al., 2004). From an evolutionary standpoint, reactions to stigma are grounded in previously-established survival mechanisms designed to protect the self from disease or threat, by avoiding association or contact with a physically or socially undesirable individual (Derosus et al., 2016; Park et al., 2007; Pryor et al., 2004; Ryan et al., 2012). In contrast, conditioned emotional reactions to stigma are learned through culturally-held associations, which become automatically engrained in individuals (Pryor et al., 2004). Regardless of the origin of the automatic association, observers viewing noticeable and undesirable features tend to fixate on those characteristics and experience immediate physiological fear or threat responses (Derosus et al., 2016). Such responses have been found to range from uncertainty, disgust, avoidance, anxiety, to engagement in stereotyping (Amodio & Devine, 2006; Derosus et al., 2016; Houston & Bull, 1994; Madera, 2016; Neumann & Strack, 2000; Pryor et al., 2004). These automatic emotional responses invoke deep visceral feelings towards stigmatized individuals, which could influence the observer's initial impressions and judgements, and ultimately colour their decisions regarding a final interview outcome.

According to the dual-process theory framework designed by Derosus et al., (2016), initial impressions and judgements are formed as a result of Type 1 emotional responses to stimuli, and are channelled into decision-making processes via cognitive scripts and heuristics. Cognitive scripts and heuristics allow individuals to quickly and automatically obtain relevant information,

and to subsequently process, interpret and integrate that information using unconscious processes (Derous et al., 2016; Evans, 2008). Such processes allow individuals to quickly determine appropriate courses of action in a given situation, while reserving working memory for more complex decision processes (Derous et al., 2016).

As such, cognitive scripts and heuristics do not require high levels of working memory or conscious decision-making to make judgements (Derous et al., 2016), however, they differ slightly in terms of how they distinguish between relevant or irrelevant information to a raters' decision. Cognitive scripts rely on innate and/or developed frameworks of knowledge or experience to "guide behaviour in familiar situations or when interacting with targets that are familiar" (Derous et al., 2016, p. 92). For example, interviewers are likely to rely on previous interview experience to guide their expectations in terms of what an appropriate decision-making process entails when rating interviews. In doing so, they may use previous experiences as a prototype to determine an appropriate course of action, even if the current situation does not exactly match the context of the prototype (Gioia & Poole, 1984). For instance, previous experience may suggest that a professional appearance is indicative of professionalism at work or strong work performance. An interviewer relying on this cognitive script in an AVI context, however, may therefore conclude that a personal background is indicative of a lack of professionalism at work or poor work performance, even though the AVI background is not exactly the same as appearance in a traditional interview setting. As such, cognitive scripts allow raters to make efficient decisions, but also potentially flawed ones given that they are applying previous situational experiences to new situations that are not necessarily identical (Gioia & Poole, 1984).

In contrast to cognitive scripts, heuristics provide “often imperfect but satisfactory reactions to certain situations or problems” (Derous et al., 2016, p. 92). In Derous et al.’s (2016) framework, heuristics can be thought of as substituting a difficult question with one that is more simple and requires less information to answer, but which also renders a less precise solution. For example, Derous et al. (2016) explain that in an interview context, the interviewer is ultimately trying to determine whether an applicant is “the optimal choice to fill the vacancy” (p. 97). In their example, answering this question would involve an in-depth analysis utilizing information which may not yet be available (i.e. interview performance scores from other candidates). Using heuristics, the interviewer instead unconsciously replaces the original question with “is this applicant right for the job” (Derous et al., 2016, p. 97), which is simpler to answer using initial information available to the interviewer. Therefore, the interviewer is able to determine the answer to the simpler question, however, the answer (or information deemed relevant to the decision) is less detailed and comprehensive that it would be to the original, more complex question. By using this less complete information to make a decision, the interviewer may introduce unconscious bias into the decision-making process.

In an interview context, Derous et al. (2016) propose that Type 1 processes are embodied by the initial reactions that interviewers experience when encountering stigmatized applicants. In other words, when interviewers notice stigmatizing features in their job applicants (or their backgrounds), they may automatically experience a negative affective response (i.e. experience feelings of avoidance or disgust), and rely on heuristics or cognitive scripts to automatically process and interpret that information. Such Type 1 processes may prompt the formation of stereotypical judgements, as the observer (or interview rater) makes rapid inferences about an individual based on limited, inaccurate or irrelevant information. As such, simply observing an

applicant's background may be enough to elicit interviewers' negative responses, and initiate the development of biases in interview ratings. Despite this concern, however, the dual-process theory also acknowledges the simultaneous involvement of more deliberate Type 2 cognitive processes in rating decisions, which are thought to temper the effects of initial Type 1 reactions (Derosus et al., 2016; Evans, 2008).

In contrast to Type 1 impulses, Type 2 processes involve conscious thought patterns that rely on working memory for functionality (Derosus et al., 2016; Evans, 2008). Type 2 processes are slower and require more resources than Type 1 processes, and result in "highly complex cognitive operations that require a high level of control and conscious awareness" (Derosus et al., 2016, p. 92). Type 2 processes also monitor Type 1 processes, and correct those Type 1 processes deemed inappropriate for a given situation (Derosus et al., 2016; Evans, 2008). According to Evans (2008), the extent to which Type 1 responses are corrected by Type 2 processes may be limited by working memory capacity, meaning that Type 2 responses may not correct for Type 1 reactions to the same extent (or at all) during times of high cognitive load or time pressure (Evans, 2008).

In the context of an interview, Type 2 processes are thought to regulate initial negative reactions to stigma, for example, by integrating professional and normative standards upheld by the workplace (Derosus et al., 2016). For instance, if the interviewer were to experience a negative initial reaction towards an individual's AVI background, the Type 2 process may incorporate conscious reasoning that such negative reactions may not be reflective of the individual's suitability for the job, or that the workplace does not condone penalizing an individual for their choice of background in a job interview. Using the more complex cognitive reasoning mechanisms inherent to Type 2 processes, the interviewers would therefore review and

adjust the inappropriate initial reaction or behaviour prompted by the Type 1 response. In cases in which the Type 1 judgement is deemed satisfactory by the Type 2 process, the individual will retain the original judgement. Despite the apparent simplicity of these two processes, there are a number of factors that influence the interactions between Type 1 and Type 2 processes throughout the course of the interview.

Initial Impression Formation

The model introduced by Derous et al. (2016) describes the interplay between Type 1 and Type 2 processes and their effects on outcomes such as initial impressions and final judgements. In doing so, the framework implicitly reflects concepts inherent to the social information processing theory that describe how initial impressions are formed. Essentially, the model suggests that initial impressions are formed when a rater “observes and interprets the explicit or implicit cues exhibited by others and makes inferences regarding their characteristics in a span of minutes” (Suen et al., 2019, p. 4). In other words, raters form initial impressions based on cues that they observe from applicants.

Given the variety of implicit and explicit visual cues that can be gleaned from candidates’ recorded backgrounds, there is arguably opportunity for interviewers to form initial impressions based on the type of background setting. Visual information has been found to be instantaneously processed by the human mind and tends to be among the first pieces of information available to others during encounters (Thorpe et al., 1996). Since initial impressions are thought to be formed in as little as the first few minutes of an interaction (Mathis et al., 2014, as cited in Torres & Gregory, 2018), the rapid processing of visual information could therefore easily inform initial impressions. In fact, several studies and reviews have supported the fact that visual cues influence initial impressions. For instance, research has shown that observers draw

inferences in terms of other individuals' personal characteristics based on physical appearances or aesthetic qualities (Nestler & Back, 2013; Suen et al., 2019; Torres & Gregory, 2018). As such, it is plausible that the visual cues found in applicants' AVI backgrounds may be responsible for generating initial impressions via Type 1 processes even before an AVI applicant speaks.

The visual cues present in applicants' AVI backgrounds may elicit immediate reactions or initial impressions from interviewers, which could bias overall interview assessments. Due to their socioeconomic status or living arrangements, it is unlikely that all applicants will have access to a private, dedicated office space in which to record their AVIs. As a result, applicants without access to an appropriate professional space may need to record their videos in other more personal spaces within their homes or in public areas, such as a bedroom, kitchen or in a public library. Despite the fact that these spaces could be as well-kept, tidy and devoid of any objects that signal personal information as professional home office spaces, by virtue of presenting a more intimate and personal space, these recording locations may be negatively perceived. For instance, a non-academic survey launched by a commercial company suggested that respondents attributed varying levels of intelligence, trustworthiness, approachability and professionalism to individuals with different objects in their backgrounds (i.e. plants, bookshelves, art, or a neutral wall; Signs.com, 2021). It could therefore be surmised that videos recorded in a personal space may incite negative initial reactions, as the applicant may be viewed as lacking professionalism or conscientiousness. In contrast, AVIs recorded in a professional office space may initially elicit positive reactions, as the visual information may signal that the individual possesses desirable qualities (i.e. conscientiousness and professionalism).

That said, with fewer visual cues available to the interviewer, blurred backgrounds would likely limit the degree of negative reactions associated with personal spaces, while also tempering the positive biases associated with professional spaces. Blurring the background would restrict the visibility of stigmatizing objects, thus likely reducing the negative reactions that raters would experience when viewing them. Likewise, when non-stigmatizing objects are blurred, the positive reactions that raters would normally associate with the background features would likely be lessened due their reduced visibility. It was therefore anticipated that professional backgrounds would be associated with the most positive initial impressions overall, followed by blurred backgrounds, and finally, with personal backgrounds being associated with the most negative initial impressions overall. As such, the following was hypothesized:

***Hypothesis 1:** Professional backgrounds will elicit the highest average initial impression scores, followed by blurred backgrounds, and personal backgrounds will elicit the lowest average initial impression scores.*

Indirect Effects of Initial Impressions on Interview Performance

In addition to informing initial impressions in an AVI, the background also provides visual cues to the rater throughout the course of the interview, thus likely contributing to a direct relationship between background setting and final interview scores. One study found that initial impressions formed during the rapport-building stage of the interview dissipated as the interview progressed (Swider et al., 2016), however, the effects of readily and consistently visible stigma were not examined. Therefore, it is possible that visible stigma that is present throughout an AVI, such as the background setting, may continuously colour the performance ratings awarded to the candidate throughout the course of the interview. As such, a direct relationship between background setting and final interview scores was anticipated, however, it was expected that this

relationship would be fully mediated by the anchoring effect associated with raters' initial impressions.

Initial impressions are not always accurate representations of candidates' suitability for a position (Eisenkraft, 2013; Torres & Gregory, 2018), and therefore they can affect the validity of an interview due to their potential anchoring effects on final interview assessments. Anchoring bias suggests that initial decisions anchor subsequent final decisions, meaning that individuals tend to avoid significant deviations from their initial assessments (Tversky & Kahneman, 1974). Although there is some evidence to suggest that initial impressions have less of an effect on final judgements in an AVI context (compared to synchronous video interviews), research has nevertheless identified that an anchoring effect persists within this modality (Suen et al., 2019).

In the present study, the relationship between the background setting and final interview scores was anticipated to be mediated by the anchoring effect of raters' initial impressions. For instance, applicants with professional backgrounds would likely elicit positive initial impressions, which would then be associated with higher final performance scores. In contrast, candidates with personal backgrounds would likely elicit negative initial impressions, which would be associated with lower final performance ratings as a result of anchoring bias. Finally, candidates using blurred backgrounds would likely elicit neither positive nor negative initial impressions, which would be associated with neutral final performance ratings. As such, it was hypothesized that:

***Hypothesis 2:** Initial impression scores will be positively associated with final interview performance scores.*

***Hypothesis 3:** Professional backgrounds will elicit the highest average final interview scores, followed by blurred backgrounds, and personal backgrounds will elicit the lowest average final interview scores.*

***Hypothesis 4:** Initial impressions will mediate the relationship between the background setting and interview performance scores.*

Feeling of Rightness

This study also examined raters' feeling of rightness as an additional exploratory area of interest. In interviews, the strength of raters' feelings of rightness is thought to represent the anchoring mechanism that either encourages or inhibits judgements from being updated once initial impressions have been formed (Derous et al., 2016). In essence, the feeling of rightness is a metacognitive intuition, or individual perception or state, which reflects the level of confidence that an individual holds regarding the accuracy of their Type 1 responses (Derous et al., 2016; Thompson et al., 2011). Feelings of rightness are stronger when the heuristic or cognitive script is readily available or easily recalled, which provides a stronger sense of being correct (Thompson et al., 2011, 2013).

Because individuals more quickly form initial impressions and experience stronger affective and physiological responses when observing a stigmatizing feature than a non-stigmatizing one, feelings of rightness are thought to be stronger when observing stigmatized individuals (Derous et al., 2016). A strong feeling of rightness subsequently prevents raters from challenging their original decisions, resulting in a lower likelihood that initial impressions will be challenged even when contradicting information is presented (Thompson et al., 2011, 2013). Derous et al.'s (2016) model therefore proposes that stigmatized candidates experience more anchoring bias in their performance scores compared to non-stigmatized candidates.

In contrast, when facing a non-stigmatized individual, an interviewer will experience a reduced physiological response and will take longer to form initial impressions, which are thought to be associated with reduced feelings of rightness (Derosus et al., 2016; Thompson et al., 2011, 2013). Reduced feelings of rightness subsequently lead individuals to engage in more conscious cognitive processes (i.e. Type 2 processes) to re-assess and challenge initial impressions (Thompson et al., 2011, 2013). For instance, if the latest interview response information is incompatible with a previously drawn heuristic-based conclusion, the Type 2 processes will subsequently update the original judgements with the new information (Derosus et al., 2016). If the latest interview response information is consistent with the initial reactions, however, then the Type 2 processes will maintain the original heuristic-based conclusion. As such, raters of non-stigmatized applicants are more likely to engage in conscious consideration of the applicant's suitability for the job, rather than relying on heuristics or cognitive scripts.

Because a high degree of subjectivity and individual differences may be at play when assessing the feeling of rightness, this topic was addressed via the following research question:

***Research Question 1:** Are high feelings of rightness associated with a stronger relationship between initial impressions and final interview scores? Are low feelings of rightness associated with weaker relationships between initial impressions and final interview scores?*

Moderating Effects

There are also a number of factors that likely moderate the relationship between initial impressions and final performance scores. Two particularly salient potential moderators are the raters' need for cognitive closure and the quality of applicant responses.

Raters' Need for Cognitive Closure

Need for cognitive closure (NFCC) is a stable individual difference that has been found to influence the extent to which raters re-evaluate their previous judgements or change their initial assessments based on new and contradicting information (Alison et al., 2008; Buijsrogge et al., 2021). NFCC represents the extent to which a rater accepts ambiguity and is willing to adjust their opinions in the face of changing circumstances or new information that conflicts with their previously-established opinions (Webster & Kruglanski, 1994). One study examining visual stigma (i.e. port-wine stains) found that NFCC moderated the extent to which raters relied on strong visual cues to make final assessments of candidates (Buijsrogge et al., 2021). In that study, NFCC influenced the extent to which raters' initial impressions formed during the rapport-building stage were influenced by job-relevant information learned during the formal interview.

Individuals high in NFCC are typically less inclined to adjust preconceived notions, and instead make judgements that rely more heavily on heuristics and stereotypes (Alison et al., 2008; Buijsrogge et al., 2021). For example, one study found that police interviewers high in NFCC tended to deviate from interview best practices and used coercion in an effort to reduce the sense of ambiguity surrounding an incident (Alison et al., 2008). This finding is consistent with the concept of Type 1 processes, in that the police interviewers sought information to confirm their impressions, rather than using more objective techniques to consider various possibilities or scenarios. In contrast, individuals low in NFCC have been found to be more likely to consider, weigh and integrate conflicting information into their decision-making processes (Alison et al., 2008; Buijsrogge et al., 2021), thereby resulting in less biased judgements. Individuals low in NFCC could therefore be thought of as relying more heavily on Type 2 processes to rationalize their judgements.

In an AVI context, individuals high in NFCC would therefore be more likely to rely on their initial impressions of a candidate, rather than the candidate's responses to the person-job fit questions (i.e. their suitability for the target position). Individuals low in NFCC, however, would be more likely to integrate the subsequent person-job fit information into their initial impressions, in order to obtain a more objective and accurate assessment of an applicant's suitability. As such, the following was hypothesized:

***Hypothesis 5:** Raters' need for cognitive closure (NFCC) will moderate the relationship between initial impressions and final interview performance scores, in that it will be stronger (versus weaker) for raters high (versus low) in NFCC.*

Quality of Candidate Responses

Response quality is also expected to play a moderating role in the relationship between initial impressions and final interview decisions. Since high quality candidates effectively answer structured interview questions designed to assess person-job fit, high quality candidates should theoretically earn strong scores that reflect their suitability for the position (Catano et al., 2022). In contrast, low quality candidates would normally receive low scores that indicate a lack of fit between their responses and the competencies needed to successfully fulfill a role (Catano et al., 2022). As previously discussed, however, the expected final interview scores may become anchored by initial impressions, which may subsequently bias final interview outcomes. A study by Roulin and Bhatnagar (2021), for instance, demonstrated that the anchoring effect of initial impressions was weaker for high quality responses compared to average quality responses. This finding suggests that raters engage in more Type 2 conscious processing when assessing high quality responses, and may rely more heavily on initial impressions and Type 1 processes when assessing lower quality responses. As such, it was hypothesized that:

Hypothesis 6: Applicant response quality will moderate the relationship between initial impressions and final interview scores, in that the relationship between initial impressions and final interview scores will be weaker (versus stronger) for high (versus low) quality candidate responses.

Methods

Design

This study was conducted using a 3 (professional background versus a personal background versus a blurred background) x2 (high quality versus low quality applicant responses) between-subjects design.

Mock interviews were pre-recorded, in which an actor played the part of a candidate applying for a remote work business operations manager position. Because individuals representing more than one stigmatized group (i.e. female, older workers, minority group members, etc), tend to experience varying levels of discrimination across different hiring contexts (i.e. job type and status), and depending on applicant and rater traits (Derous et al., 2012; Derous et al., 2015; Ruggs et al., 2014), the mock interview actor was selected based on personal characteristics that limited the introduction of additional stigma to the greatest extent possible. As such, the actor was a Caucasian male, in the hopes of reducing the amount of additional stigmatizing features associated with the “AVI candidate.” Given that the target position was not an entry-level role, the actor was in his 30s to make the scenario as a realistic as possible.

A remote work position was selected for the target job, as it was anticipated that an individual’s background may be more salient if the applicant is expected to use video modalities for their future work, following selection. In other words, the applicants’ backgrounds may be

perceived as a preview of the work environment in which they would engage with clients or colleagues in remote work contexts. An AVI background may therefore likely be perceived as less relevant to positions targeting traditional office environments, as the applicant would ultimately conduct their business in the formal office space. As such, the business operations manager job description was adapted from a previous student's MSc thesis (Wong, 2020) to reflect a remote work context (Appendix A).

The actor recorded AVI responses using either a high or low quality response script, to reflect either strong or weak responses to the five structured interview questions (Appendix B). The responses were designed to elicit either scores ranging from 1 to 2 (low quality responses) or scores ranging from 4 to 5 (high quality responses), out of a maximum possible 5 points per question. The target score for each response was derived from the Behavioural Anchored Rating Scales (BARS) developed in a previous student's MSc thesis (Wong, 2020). The high and low quality response scripts for each question were similar in content to the greatest extent possible.

In addition to the interview question response videos, the actor also recorded introductory videos for each of the manipulated conditions. These videos simulated the practice video opportunities offered by AVI platforms and doubled as a safeguard for the mock organization to confirm applicants' tombstone data, which could be plausibly occur in a real AVI context. This introductory video was used to assess initial impressions in the absence of a rapport-building phase of the interview. During this introductory video, the actor stated his name (an alias), the position being sought, and the competition number. This introductory video therefore provided raters with an opportunity to record initial impressions while limiting the confounding effects of ancillary information.

The background in each video was manipulated to reflect different types of background setting (i.e. professional versus personal versus blurred; see Appendix C). The professional background portrayed a tidy and neutral private office space (i.e. bookshelf, lamp, framed diplomas, and plants), and the personal background depicted a tidy and neutral bedroom (i.e. bed, nightstands, framed artwork, and lamps). The blurred background manipulation was achieved by recording the videos using the blurred settings available on Zoom. To ensure consistency across conditions, each interview question by condition was recorded twice, and subsequently compared to the other conditions. The recorded videos were compared in terms of the quality of the video's lighting and audio, as well as the frequency of the actor's eye contact and non-verbal behaviours (i.e. gestures, nodding, smiling).

Pilot Studies

Pilot Study 1A. An initial pilot study (Pilot Study 1A) was first conducted in order to ensure the effectiveness of the manipulations. A group of six graduate-level students (male: $N = 4$; female: $N = 2$) belonging to the Psychology Department at a Canadian university was asked to review a job description for clarity and understanding, and to assess either the high- or low-quality response scripts for each question (i.e. each participant reviewed 2-3 high quality response scripts and 2-3 low quality response scripts) to confirm that the scripts elicit high or low ratings as anticipated. The students were also presented with still photos of the selected background locations, in order to gauge whether they were initially perceived as professional versus personal backgrounds. Based on the feedback received from pilot study participants, some of the response scripts were edited so that they would more consistently reflect either higher or lower quality responses.

Response script ratings. Participants were asked to rate the quality of the response scripts and the rating scale options ranged from 1 (“very poor response”) to 5 (“very good response”). A bootstrapped one-tailed paired samples t-test was used to analyse the Pilot Study 1A data, and indicated that there were significant differences between the high and low quality response script ratings, $t(5) = 6.41, p < .001, 95\% \text{ CI}[1.44, 3.39]$, which represented a very large effect ($d = 3.99$). On average, participants evaluated the high-quality response scripts as being representative of higher quality responses ($M = 4.56, SD = 0.55$) compared to the low-quality response scripts ($M = 2.14, SD = 0.65$). Closer examination of individual items and comments (as shown in Table 1), however, indicated that some revisions were required to more clearly differentiate between the high- and low-quality response scripts in some cases. The low-quality response scripts for Questions #1, 2, and 4, as well as the high-quality response script for Question #5 were therefore revised.

Still photo professionalism ratings. Participants were asked to rate the professionalism of the professional and personal backgrounds visible in still photos, and the rating scale options ranged from 1 (“very unprofessional”) to 5 (“very professional”). A bootstrapped one-tailed paired samples t-test also indicated that there were significant differences between the still photo background professionalism ratings, $t(5) = 13.00, p < .001, 95\% \text{ CI}[1.74, 2.60]$, which represented a very large effect ($d = 4.65$). On average, Pilot Study 1A participants rated the professional background as being more professional ($M = 4.33, SD = 0.51$) than the personal background ($M = 2.17, SD = 0.41$).

Pilot study 1B. Following the response script edits, the same pilot study process was repeated with a second group of graduate students (Pilot Study 1B) to confirm that the response scripts elicited the anticipated scores. Seven graduate students belonging to the Psychology

Department at a Canadian university participated in Pilot Study 1B. One participant in Pilot Study 1B demonstrated a misunderstanding of the pilot study instructions, which necessitated the removal of that individual's data, leaving a total of $N = 6$ participants in the Pilot Study 1B (male: $N = 3$; female: $N = 3$).

Response script ratings. Another bootstrapped one-tailed paired samples t-test indicated that again there were significant differences between the high and low quality response ratings, $t(5) = 17.33, p < .001, 95\% \text{ CI}[2.74, 3.70]$, which represented a very large effect ($d = 9.71$). On average, participants evaluated the high-quality response scripts as being representative of higher quality responses ($M = 4.53, SD = 0.40$), compared to the low-quality response scripts ($M = 1.31, SD = 0.25$). Closer examination of the individual ratings provided by participants during Pilot Study 1B (see Table 1) showed that participants were able to more consistently discern high- and low-quality responses.

Still photo professionalism ratings. Another bootstrapped one-tailed paired samples t-test indicated that there were significant differences between the still photo background professionalism ratings, $t(5) = 4.47, p = .003, 95\% \text{ CI} [.85, 3.15]$, which represented a very large effect ($d = 2.83$). On average, Pilot Study 1B participants rated the professional background as being more professional ($M = 4.50, SD = 0.55$) than the personal background ($M = 2.50, SD = 0.84$).

Table 1*Response Script and Still Photo Ratings for Pilot Studies 1A and 1B*

Question No.	Response Quality	Pilot Study 1A			Pilot Study 1B		
		<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
1	High	3	5.00	0.00	3	3.67 ^a	1.53
1	Low	3	2.67	0.58	3	1.00	0.00
2	High	4	5.00	0.00	2	5.00	0.00
2	Low	2	2.50	0.71	4	1.75	0.50
3	High	3	4.67	0.58	3	4.67	0.58
3	Low	3	1.00	0.00	3	1.33	0.58
4	High	2	5.00	0.00	4	4.75	0.50
4	Low	4	3.00	1.83	2	1.00	0.00
5	High	2	3.00	1.41	4	4.75	0.50
5	Low	4	1.50	0.58	2	1.00	0.00
Still Photo							
Background							
	Professional	6	4.33	0.52	6	4.50	0.55
	Personal	6	2.17	0.41	6	2.50	0.84

Note. ^aOne participant awarded a lower score for this item, but indicated in the comments section that the reason was unrelated to the quality of the response. The response script was therefore retained without further edits.

Pilot Study 2. Next, Pilot Study 2 was conducted via the MTurk online platform in order to confirm whether participants were able to distinguish between blurred professional and personal backgrounds, using high quality candidate response videos. This pilot study was primarily conducted due to the uncertainty surrounding whether or not individuals were able to distinguish between a blurred professional or blurred personal setting. MTurk workers were therefore asked to rate the extent to which they were able to identify items in the background of the blurred conditions, and were asked to note the objects that they recognized from a predetermined list. Participants were also asked to rate the level of professionalism that they

associated with the blurred professional and personal backgrounds presented in the videos, as well in still photos of the visible professional and personal backgrounds.

Pilot Study 2 participants consisted of 115 anglophone MTurk workers (male: $N = 70$; female: $N = 43$; non-binary: $N = 1$). This sample size was selected based on an a priori G*Power power analysis using a one-tailed t-test (difference between two independent means) with a medium effect size of 0.50 (power = .80).¹ The recommended sample size of 102 was increased to a sample size of $N = 115$ to account for any missing or poor-quality data. Participant age ranged from 23 to 70 years old ($M = 36.54$, $SD = 9.86$), and most (47.8%) held a bachelor's degree (high school diploma: 29.6%, Master's degree: 10.4%, Associate's degree: 6.1%, PhD: 2.6%, College diploma: 1.7%, less than a high school diploma/GED: 0.9%, no response: 0.9%). The majority of participants (80.9%) were Caucasian (Black/African-American: 13.0%; Hispanic/Latino: 8.7%; Asian: 5.2%; Mixed Race: 1.7%; Middle Eastern: 0.9%; Other: 0.9%), and despite the study being available to participants residing in Canada or the United, all of the participants reported living in the United States (no response: $N = 1$).

Assumptions. The normality of the Pilot Study 2 data distribution was examined by visually inspecting histograms and normal Q-Q plots, verifying the skew, kurtosis data, and checking for outliers using standardized score distributions, and examining the outcomes of Shapiro-Wilk tests. The difference between the blurred professional and personal perceived object visibility rating data and the difference between the still photo professionalism rating data were normally distributed, despite the outcomes of their respective Shapiro-Wilk tests (perceived

¹ I was initially planning to use a between-person design in the pilot study, which is reflected in the power analysis results. Although a smaller sample would have been sufficient for the within-person design that was ultimately used, the original sample size was retained in order to more confidently determine whether or not objects were visible in the blurred conditions.

object visibility ratings: $W(114) = 0.89, p < .001$; still photo professionalism ratings: $W(115) = 0.92, p < .001$). The difference between the blurred professional and personal accuracy score data were platykurtic, but otherwise normally distributed. The difference in video professionalism rating data across the two conditions were leptokurtic, negatively skewed, and showed evidence of violations of normality based on inspection of the histogram, Q-Q plot and Shapiro-Wilk test, $W(115) = 0.80, p < .001$, and therefore bootstrapping was used for analysis of this data.

Perceived ability to identify objects. First, the Pilot Study 2 data was analyzed to determine how well participants perceived that they were able to identify objects in the background of the blurred professional and personal videos. On average, participants reported being unable to identify most items in the background of both the blurred professional video ($M = 2.37, SD = 0.76$), and the blurred personal video conditions ($M = 1.66, SD = 0.83$). The average difference between the object visibility ratings in the blurred professional and personal conditions was 0.71 ($SD = 0.91$). A paired samples t-test indicated that participants perceived that they were able to identify significantly more background objects in the blurred professional background condition compared to the blurred personal background condition, $t(113) = 8.34, p < .001, 95\% CI [.54, .88]$, which represented a large effect ($d = 0.89$).

Object Identification Accuracy. Despite rating their ability to recognize objects as higher in the blurred professional background condition, participants were not significantly more accurate in identifying particular items viewed in the videos compared to in the blurred personal condition. Participants were provided with a list of items that were truly present in the video background as well as some objects that were not included in the background, and were asked to identify all of the objects that were visible in the respective condition (i.e. blurred professional or blurred personal background). Object identification accuracy scores were calculated by

determining the difference between the number of correctly identified objects and incorrectly identified objects in each condition. Participants did not demonstrate a notable ability to correctly select particular objects visible in the background nor to avoid selecting those that were not present in the video backgrounds in either condition. The average difference between the accuracy scores in the blurred professional and personal conditions was -0.14 ($SD = 1.35$). The results of a two-tailed paired samples t-test indicated that there was no significant difference between the two blurred video conditions in terms of the participants' ability to select the correct background items from a list, $t(13) = -.40$, $p = .70$, 95% CI $[-.92, .64]$, which represented a small effect ($d = 0.13$).²

Blurred Video Background Professionalism Ratings. Pilot Study 2 also re-evaluated the professionalism ratings that participants assigned to the blurred professional and personal video conditions. On average, the blurred videos were not viewed differently in terms of professionalism across the two conditions (blurred professional condition: $M = 3.30$, $SD = 1.02$; blurred personal condition: $M = 3.43$, $SD = 1.09$). The results of a bootstrapped two-tailed paired samples t-test indicated that there were no significant differences in the professionalism ratings assigned by the two conditions, $t(114) = -1.68$, $p = .100$, 95% CI $[-.30, .03]$, which represented a small effect ($d = 0.12$).

Still Photo Background Professionalism Ratings. In order to confirm the findings from Pilot Studies 1A and 1B using a larger sample, Pilot Study 2 also asked participants to rate the professionalism associated with the visible professional and personal background settings using still photographs. On average, participants rated the visible professional background as

² Bivariate correlations indicated that there was no discernable relationship between perceptions of and ability to identify objects (blurred professional condition: $r = .18$; blurred personal setting: $r = -.12$).

“professional” ($M = 4.17$, $SD = 0.75$), and the visible personal background as “neither professional nor unprofessional” ($M = 2.57$, $SD = 1.02$). The average difference between professionalism ratings across the two conditions was 1.60 ($SD = 1.07$). A one-tailed paired samples t-test indicated that participants rated the professional background as significantly more professional than the personal background, $t(114) = -15.97$, $p < .001$, 95% CI[-1.80, -1.40], which represented a very large effect ($d = 1.79$).

Order Effects. The respective results of one-way repeated measures ANOVAs indicated that the order in which the conditions were presented to participants had no significant effect on blurred video professionalism ratings, $F(1, 113) = 0.21$, $p = .644$, partial $\eta^2 = .002$, on object identification accuracy scores, $F(1, 12) = 2.65$, $p = .129$, partial $\eta^2 = .181$, nor on still photo background professionalism ratings, $F(1, 113) = .02$, $p = .890$, partial $\eta^2 = .000$. The results of a one-way repeated measures ANOVA did, however, indicate the presence of a significant interaction between the condition order and participants’ ratings of their perceived ability to identify objects in the video backgrounds, $F(1, 112) = 19.30$, $p < .001$, partial $\eta^2 = .147$. Perceived item visibility in the professional background condition was rated as being significantly higher ($M = 2.60$, $SE = 0.09$) when the blurred personal background video was presented before the blurred professional background video, $F(1, 112) = 12.59$, $p < .001$, compared to when the order was reversed ($M = 2.13$, $SE = 0.10$). There was no order effect on participants’ ratings of their perceived ability to identify objects in the personal background condition, $F(1, 112) = 19.30$, $p = .165$.

Choice of blurred video to be used in the main study. The blurred personal videos were selected over the blurred professional videos to represent a “neutral” background condition in the main study. The decision to use the blurred personal videos was made based on three factors.

First, the Pilot Study 2 results indicated the presence of order effects in the blurred professional background video condition, and given the lack of clear explanation as to why these effects occurred, using the blurred professional videos risked contributing additional confounding effects to the main study. Second, Pilot Study 2 participants rated their perceived ability to identify objects in the background as being higher in the blurred professional video condition. Therefore, the use of the blurred professional video condition in the main study could introduce additional sources of bias based on participants' perceptions of what they see in the video backgrounds. Finally, in a practical AVI context, applicants are logically more likely to blur their backgrounds in an attempt to hide personal spaces rather than professional ones in order to improve the outcomes of their interviews. As such, the blurred personal videos were selected to represent the "neutral" background setting condition in the subsequent main study.

Procedure

The hypotheses³ and conceptual model of the study design (see Figure 1) were adjusted following the analysis of the Pilot Study 2 data, and the main study was subsequently launched using the Prolific platform. Participants were advised that the study would examine how AVI features impact raters' initial impressions and applicants' final interview scores, and that they were required to rate structured AVI responses. Participants' informed consent was obtained prior to their random assignment to one of the six conditions.

³ The original study design supposed that the blurred professional and personal backgrounds were distinguishable from one another, and that the visibility of the AVI background (i.e. clearly visible versus blurred) moderated the relationship between background setting and initial impressions. Based on the results of the pilot study (the two blurred backgrounds were in fact indistinguishable from one another), the blurred background was incorporated as a third "neutral setting" condition to the background setting manipulation, and the moderating role of background visibility was removed from the design.

To commence the study, participants were first provided with instructions as to what an AVI entails, and how to score the recorded interview responses. They were then asked to review the job description, before being shown the candidate introductory video. Participants were then asked to evaluate their initial impressions of the candidate, before moving on to the formally scored portion of the interview. Five recorded video responses (corresponding to one of the six manipulated conditions) were presented to participants in the same order across conditions. Raters had the opportunity to watch the recorded videos as many times as required, but were required to watch the entirety of each video prior to moving on to the next interview response. To simulate a real AVI assessment, participants were asked to assign a score (i.e. a number of stars) following each recorded response. Once all of the five videos had been assessed, participants were asked to rate the candidate's overall final interview performance. Following the final interview performance ratings, participants were asked to rate the strength of their feeling of rightness, complete an NFCC scale, and answer demographic questions. Finally, participants were debriefed in terms of the study's intent to study the potential biasing effects of background setting across varying quality of responses, and were provided with contact information should they wish to provide feedback to the researcher.

Participants

Participants of the main study were 400 anglophone Prolific users residing in Canada ($n = 110$) and the United States ($n = 288$; no response: $n = 2$). This sample size was calculated in G*Power by first estimating the a priori sample size needed for each main effect (i.e. background setting, background visibility, initial impressions, NFCC, and response quality). The sample sizes of each interaction represented in the conceptual model were also estimated (i.e. initial impressions/NFCC, and initial impressions/response quality; UCLA: Statistical Consulting

Group, 2021). Each respective effect size was estimated using variance values previously reported in the literature (Alison et al., 2008; Roulin & Bhatnagar, 2021), apart from the background setting variable. Since to the best of my knowledge, background setting has not yet been examined in previous studies, it was not possible to estimate the precise values of its main effect and interaction. The remaining estimations ranged from sample sizes of $N = 178$ to $N = 366$, and the largest sample size ($N = 366$) was selected to enhance the likelihood that indirect effects would be detected in the study (Preacher et al., 2007). In order to account for the possibility of poor quality or missing data, this sample size was increased to $N = 400$.

Approximately half of the participants were female (49.5%, $n = 198$; male: 48.5%, $n = 194$; other: 2.0%, $n = 8$), and their ages ranged from 18 to 81 years old ($M = 43.28$, $SD = 13.75$). Most participants held a Bachelor's degree (43.8%; Master's degree: 21.8%; High School diploma/GED: 15.5%; Associate's degree: 8.3%; College diploma: 6.5%, PhD: 4.0%). The majority of participants reported being White/Caucasian ($n = 321$; Asian: $n = 39$; Black/African-American: $n = 25$; Mixed race: $n = 10$; Hispanic/Latino: $n = 18$; Middle Eastern: $n = 9$; Native/Aboriginal/Indigenous: $n = 5$; Prefer not to say: $n = 1$).

Most respondents reported being currently employed (81.8%, $n = 327$; no response: $n = 3$), and the majority worked full-time (61.3%, $n = 245$; self-employed/independent: $n = 47$; part-time: $n = 41$; unemployed: $n = 32$; other: $n = 30$; contract work: $n = 3$; casual/seasonal: $n = 2$) for an average of 34.10 hours per week ($SD = 15.72$). On average, participants had worked at their current job for seven years ($M = 6.84$, $SD = 7.81$). Approximately half of the participants reported currently working remotely (52.0%, $n = 208$), and reported having worked remotely for 2.54 years ($SD = 2.52$). Many participants reported working remotely in the past (47.8%, $n = 191$; no response: $n = 1$), and on average reported working remotely in the past for 2.88 years

($SD = 3.74$). Approximately one-third of participants reported working remotely due to existing COVID-19 circumstances (35.6%, $n = 145$), and reported having done so for the past 1.77 years on average ($SD = 0.90$). The majority of participants indicated that they had a private, dedicated workspace in their homes (67.0%, $n = 268$).

Most participants had conducted face-to-face interviews (96.3%, $n = 385$), and reported having conducted an average of 36 interviews in the past ($M = 35.68$, $SD = 80.35$, median = 15.0, mode = 10). Most participants had never participated in an AVI as either an applicant (80.8%, $n = 323$) or as a rater (82.0%, $n = 328$). The majority of respondents reported rating between one and 30 AVI interviews ($M = 13.53$, $SD = 42.42$; median = 5.00; mode = 2.00), however, two participants reported rating several hundred. Approximately half of participants had previously conducted a video conference interview (49.8%, $n = 199$), and estimated conducting an average of 17 video conference interviews in the past ($M = 17.28$, $SD = 72.57$, median = 5.0, mode = 5.0).

Measures

Initial impression and final interview scores. Initial impressions were assessed using four items adapted from scales previously used to globally assess applicant suitability (Cable & Judge, 1997; Higgins & Judge, 2004; Stevens & Kristof, 1995). The initial impression scale showed good reliability ($\alpha = .94$). Items were rated on a scale from 1 (“Not at all/Extremely low”) to 7 (“Extremely/Extremely high”), and included: “How qualified is this applicant for the job?,” “How attractive is this applicant as a potential employee?,” “How highly do you regard this applicant?” and “How highly does this applicant fit the target position?.” The average scores across the first set of the four items (completed after the introductory video) represented the initial impression score, and the average score across the second set of items (completed

following the interview response videos) represented the final interview performance score. The final interview score scale also showed good reliability ($\alpha = .99$).

Response-level scores. To simulate a real AVI rating context, participants were asked to rate each of the five interview response videos. Participants were asked to “Please assign a number of stars to this interview response” and response options ranged from 1 (“Low”) to 5 (“High”). This item was intentionally designed to be vague, in order to avoid priming candidates towards more objective measures of suitability (i.e. standardized scales), which could potentially reduce the biasing effects of the background and visibility manipulation. The results from the five interview questions were aggregated, and the average score represented the average individual response scores. The average individual response scores showed good reliability ($\alpha = .97$).

Need for cognitive closure (NFCC). Raters’ NFCC was assessed using the 15-item Need for Closure Scale (NFCS; Roets & Van Hiel, 2007, 2011; Webster & Kruglanski, 1994). The NFCS assesses the extent to which an individual desires certainty, dislike ambiguity, as well as forms quick impressions and strong opinions. This scale was found to have good reliability ($\alpha = .89$). Response options ranged from 1 (“Strongly disagree”) to 6 (“Strongly agree”). Scores of up to 30 points indicated low NFCC, and scores between 75 and 90 suggested high NFCC (Kruglanski, 2021).

Feeling of rightness. Feeling of rightness was evaluated using three items adapted from Buijsrogge et al. (2016), which included “I’m confident in the interview ratings that I assigned,” “I believe that I successfully evaluated the interview,” and “I am satisfied with my interview assessments.” A 5-point scale was used, ranging from 1 (“Absolutely disagree”) to 5 (“Absolutely agree”). The scale showed good reliability ($\alpha = .89$).

Demographics. Participants were asked to indicate their sex, age, level of education, ethnicity, years of work experience, years of experience evaluating interview responses, the number of interviews that they have conducted in the past, the number of AVIs that they had rated or completed, years of remote work experience, and whether they had a private, dedicated workspace in their homes.

Data quality checks. An attention check item was embedded among the initial impression items (“Select ‘Not at all’ for this item”), as well as among the NFCS items (“I enjoy eating concrete”). In addition, participants were asked whether they considered their data to be of high quality (“yes/no”), and if their data should be used for the study (“yes/no”).

Manipulation check. Candidates were asked to rate the level of professionalism of the recorded video background on a scale ranging from 1 (“Very unprofessional”) to 5 (“Very professional”). Participants were also asked “Did the applicant in the recorded videos use a background blurring setting,” with a “Yes/No” response option.

Awareness of the study manipulations. As a final measure in the study, participants were asked to indicate what they believed the study focus was in open-text response form.

Results

Assumptions. In order to rule out any potential sources of bias that may influence the regression outcomes, I confirmed whether or not the data respected the assumptions of the linear model. To do so, I inspected all of the variables for missing or abnormal data via frequency tables, verified that all data points fell within the expected scale ranges, and confirmed that the amount of missing data never exceeded 5% of the total sample size (Meyers et al., 2017). Apart from the NFCC composite scale, all of the scales used in the analyses showed violations of

normality based on visual inspection of the histograms and normal Q-Q plots, skew and kurtosis data, and examination of the Shapiro-Wilk test outcomes. Examination of Cook's distance scores indicated that there were no influential cases, however, inspection of Mahalanobis distances pointed to three multivariate outliers (critical value = 20.515, $p = .001$, $df = 5$). The three multivariate outliers were retained in the dataset, as their removal did not alter the overall results of the study. Scatterplots of the study variables showed acceptable linearity. I further screened the study variables within the three background setting conditions and within the two response quality levels respectively, and found that the composite scales showed similar patterns of normality across all of the background setting conditions. The composite scales also showed similar patterns of normality across all of the response quality conditions, with the exception of the composite response-level score data, which was more skewed and kurtotic in the low-quality response condition. There were equal variances across the background setting conditions, however, the Levene's test (based on the mean) indicated that there were equal variances across the response quality conditions only for some scales (i.e. initial impressions and NFCC), but not for others (i.e. response-level scores, final interview rating scores, and feeling of rightness). The Box's test also indicated that there was inequality of variance-covariance across the levels of both background setting and response quality. Based on these violations of the linear model assumptions, the subsequent analyses were bootstrapped in order to offset their potentially biasing effects.

Manipulation checks. On average, participants rated the professional background as being more professional ($M = 4.08$, $SD = 0.80$) than the personal ($M = 3.68$, $SD = 0.89$) or the

blurred ($M = 3.67$, $SD = 0.81$) backgrounds.⁴ Most participants correctly identified whether their assigned condition had a blurred background or not (professional background condition: 98.5%, personal background condition: 98.5%; blurred condition: 93.5%).⁵

Attention checks. All participants passed at least one of the two attention checks, and therefore all data was retained within the dataset.

Data quality checks. One participant indicated that their data was of poor quality and should not be used within the study. Given that removal of that participant's data did not change the outcome of the analyses, the data was retained within the dataset.

Descriptive Statistics and Intercorrelations

Descriptive statistics and intercorrelations for all study variables are presented in Table 2. The effect sizes were interpreted using the interview behaviour correlational effect size benchmarks described by Bosco et al. (2015), which were identified as being more appropriate standards to use when conducting research in applied psychology. Bosco et al. (2015) separated the interview behaviour effect size distribution into four equal parts using percentiles. They suggest that effect sizes less than $r < .08$ fall between the 1st and 25th percentiles of the

⁴ A bootstrapped, one-way between-subjects ANOVA was conducted to confirm whether the background setting conditions were rated differently in terms of how professional participants perceived them to be. The background professionalism ratings significantly differed across the conditions, $F(2) = 10.71$, $p < .001$, and Bonferoni post-hoc comparisons indicated that the professional background was rated as significantly higher than the personal ($p < .001$, 95% CI[.20, .60]) and blurred conditions ($p < .001$, 95% CI[.24, .61]), however, there was no significant difference in professionalism ratings between the personal and blurred conditions ($p = 1.0$, 95% CI[-.20, .21]).

⁵ I conducted an additional exploratory analysis to determine if excluding the data of those who had failed the first manipulation check (identification of the blurred background) affected the overall results of the study due to a potential lack of attention. The analyses indicated that removing the participants who failed the first manipulation check did not alter the overall outcome of the study. Based on this outcome, and to avoid introducing additional non-random, differential attrition into the study (Aronow et al., 2019; Montgomery et al., 2018), the data of those who had failed the attention check was therefore retained in the study.

distribution, $r = .08$ to $.14$ fall between the 25th and the 50th percentiles, $r = .14$ to $.33$ fall between the 50th and 75th percentiles, and $r > .33$ corresponds with percentiles greater than the 75th percentile in the distribution.

Intercorrelations. The intercorrelations pointed to strong, positive relationships between response quality and response-level scores and final interview scores ($r = .871$ and $.881$, respectively, $p < .01$). The dummy-coded background setting conditions shared almost no relationship with initial impression scores ($r = -.059$ and $-.008$), response-level scores ($r = .000$ and $-.016$) or final interview scores ($r = -.010$ and $.015$). Response-level scores were highly and positively correlated with final interview scores ($r = .961$, $p < .01$), and initial impression scores shared a moderate, positive and significant relationship with final interview scores ($r = .155$, $p < .01$). There was also almost no relationship between feeling of rightness ratings and NFCC scores ($r = .035$). NFCC was weakly and positively associated with initial impression scores ($r = .091$), yet shared significant, positive, and moderately-low relationships with response-level and final interview scores ($r = .109$ and $.120$, respectively, $p < .05$). Compared to NFCC, feeling of rightness scores showed stronger negative relationships with response-level and final interview scores ($r = -.195$ and $-.184$, respectively, $p < .01$), meaning that lower response-level and final interview scores were associated with higher feeling of rightness ratings. The moderately high negative correlation between feeling of rightness and response quality ($r = -.207$, $p < .01$) indicated that low quality responses were associated with higher feelings of rightness. Feeling of rightness ratings showed almost no relationship with initial impression scores ($r = -.02$).

Descriptives. The average initial impression, response-level, final interview, feeling of rightness and NFCC ratings across the manipulated conditions are presented in Table 3. The potential for Type 1 error associated with conducting several ANOVAs across the background

setting conditions was mitigated by using Bonferoni corrections. To compensate for the number of ANOVAs being conducted, the $p = .05$ alpha level was divided by the number of dependent variables (5), as recommended by Meyers et al. (2017). A significance level of $p = .01$ was therefore used as the alpha threshold for the analyses across the background setting conditions.

Initial impression scores. On average, participants across all six conditions reported positive initial impressions ($M = 3.90$, $SD = 1.03$) towards the AVI applicant. A bootstrapped between-subjects ANOVA with Bonferroni correction showed that the initial impression scores did not significantly differ between the three background setting conditions, $F(2, 397) = 1.06$, $p = .349$.

Response-level scores. A bootstrapped between-subjects ANOVA showed that participants assigned to the high quality response conditions assigned significantly higher response-level ($M = 4.25$, $SD = 0.54$) scores compared to those assigned to the low quality response ($M = 1.74$, $SD = 0.84$) conditions, $F(1, 391) = 1220.31$, $p < .001$. The response-level scores did not significantly differ between the three background setting conditions, $F(2, 397) = 0.073$, $p = .930$. A factorial ANOVA between the background setting and response quality conditions indicated a lack of significant interaction on the response-level outcome, $F(2, 394) = 0.508$, $p = .602$.

Final interview ratings. A bootstrapped between-subjects ANOVA showed that participants assigned to the high quality response conditions assigned significantly higher final interview ($M = 5.87$, $SD = 0.97$) scores compared to those assigned to the low quality response conditions ($M = 1.80$, $SD = 1.22$), $F(1, 391) = 1631.89$, $p < .001$. The final interview scores did not significantly differ between the three background setting conditions, $F(2, 397) = 0.123$, $p = .884$. A factorial ANOVA between the background setting and response quality conditions

indicated a lack of significant interaction on the final interview score outcome, $F(2, 394) = 0.360, p = .698$.

Feeling of rightness. A bootstrapped between-subjects ANOVA showed that participants assigned to the low quality response conditions ($M = 4.75, SD = 0.45$) assigned significantly higher feeling of rightness scores compared to those assigned to the high quality response ($M = 4.56, SD = 0.52$) conditions, $F(1, 391) = 15.54, p < .001$. The feeling of rightness scores did not significantly differ between the three background setting conditions, $F(2, 397) = 0.411, p = .663$. A factorial ANOVA between the background setting and response quality conditions indicated a lack of significant interaction on the feeling of rightness outcome, $F(2, 394) = 0.903, p = .406$.

NFCC. A bootstrapped between-subjects ANOVA showed that participants assigned to the high quality response conditions ($M = 3.90, SD = 0.77$) assigned significantly higher NFCC scores compared to those assigned to the low quality response ($M = 3.72, SD = 0.85$) conditions, $F(1, 391) = 4.58, p = .033$. The NFCC scores did not significantly differ between the three background setting conditions, $F(2, 397) = 0.110, p = .896$. A factorial ANOVA between the background setting and response quality conditions indicated a lack of significant interaction on the NFCC outcome, $F(2, 394) = 2.30, p = .102$.

Table 2*Bootstrapped descriptive statistics and intercorrelations*

Variable	<i>M (SD)</i>	1	2	3	4	5	6	7	8	9	10
1. Response quality ^a		-									
2. Initial impression scores ^b	3.90 (1.03)	.052	.94								
3. Response-level scores ^b	3.00 (1.44)	.871**	.114*	.97							
4. Final interview scores ^b	3.82 (2.32)	.881**	.155**	.961**	.99						
5. NFCC ^c	3.82 (0.81)	.106	.091	.109*	.120*	.89					
6. Feeling of rightness ^d	4.65 (0.50)	-.207**	-.020	-.195**	-.184**	.035	.89				
7. Professional background ^e		.016	-.059	.000	-.010	.021	-.035	-			
8. Blurred background ^e		-.016	-.008	-.016	-.015	-.020	.042	-.498**	-		
9. Gender ^f		.047	-.042	.039	.025	.085	.013	.028	-.110	-	
10. Age	43.28 (13.75)	-.014	-.057	-.085	-.076	-.114	.126	-.065	.072	-.045	-

Note. $N = 400$. Cronbach's alpha (α) scores are shown in bolded text on the diagonal. * $p < .05$, ** $p < .01$

^a Represents the high and low quality response manipulation. High quality responses were labelled with a value of 1, and low quality responses were labelled with a value of 0.

^b Higher values are associated with higher/more positive ratings.

^c Higher values indicate higher levels of NFCC

^d Higher values indicate stronger feelings of rightness.

^e Represents the dummy coded background setting manipulation.

^f Males were labelled with a value of 1, females with a value of 2, and “Other” was labelled with a value of 3.

Table 3

Average initial impression, response-level, final interview, feeling of rightness, and NFCC ratings across conditions.

Response quality	Background setting	<i>n</i>	Initial impression ratings		Response-level ratings		Final interview ratings		Feeling of rightness ratings		NFCC	
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
High	Professional	68	3.90	1.09	4.26	0.59	5.78	1.05	4.56	0.57	3.90	0.72
High	Personal	67	4.01	0.92	4.24	0.47	5.88	0.83	4.55	0.50	3.79	0.74
High	Blurred	65	3.95	0.93	4.26	0.57	5.92	0.99	4.53	0.53	3.99	0.84
Low	Professional	65	3.73	1.07	1.68	0.80	1.71	1.13	4.69	0.52	3.76	0.89
Low	Personal	67	3.98	1.12	1.83	0.93	1.92	1.39	4.74	0.43	3.82	0.80
Low	Blurred	68	3.83	1.03	1.72	0.80	1.72	1.12	4.82	0.37	3.60	0.86
High ^a		200			4.25	0.54	5.86	0.96	4.55	0.53	3.90	0.77
Low ^a		200			1.74	0.84	1.79	1.22	4.75	0.45	3.72	0.85
	Professional ^b	133	3.81	1.08	2.99	1.47	3.79	2.32	4.63	0.55	3.83	0.81
	Personal ^b	134	4.00	1.03	3.03	1.41	3.90	2.29	4.65	0.48	3.81	0.76
	Blurred ^b	133	3.91	0.95	2.96	1.45	3.77	2.35	4.68	0.48	3.79	0.87

Note. Higher values represent higher levels of each variable.

^a Average scores across the high or low quality conditions without taking the background setting conditions into account.

^b Average scores across the background setting conditions, without taking the response quality conditions into account.

Direct & Indirect Effects

Linear regression and the IBM SPSS Statistics PROCESSv4.0 Macro (Model 16; Hayes, 2017) were used to test whether the effect of background setting on final interview ratings was mediated by initial impression ratings. The PROCESS Model 16 template matched the conceptual model used in the study (see Figures 1 and 2), as it represented a similar layout of direct and indirect effects, and included two moderating variables that influenced the relationship between the mediator and the dependent variable. The study's independent variable was the background setting manipulation, which was identified as a multicategorical variable using the indicator coding system, with the personal background condition used as a reference group. Initial impression scores represented the mediating variable, and final interview scores represented the dependent variable in the study. Applicant response quality and NFCC were the moderating variables, which were expected to interact with initial impression scores to influence the final interview scores. The analyses were bootstrapped using 5000 samples, and the continuous variables defining products were mean centered. The unstandardized regression weights and *SEs* of the model are depicted in Figure 2.

The 'a' path of background setting on initial impression scores was tested using a bootstrapped regression analysis. Overall, background setting accounted for 0.53% of the variance in initial impression scores, and the model was non-significant ($R^2 = .005$, $F[2, 397] = 1.06$, $p = .349$). Neither the professional ($b = -0.182$, $SE = 0.126$, $p = .150$, bootstrapped 95% CI[-.43, .07]) nor the blurred background ($b = -0.108$, $SE = 0.126$, $p = .390$, bootstrapped 95% CI[-.36, .14]) conditions were associated with higher initial impression scores compared to the personal background condition. Hypothesis 1 was therefore not supported.

The 'b' path was assessed by testing whether initial impression scores were a significant predictor of final interview scores, while controlling for background setting. The results indicated that initial impression scores were positively and significantly associated with final interview scores ($b = 0.263$, $SE = 0.071$, $p < .001$, bootstrapped 95% CI[.12, .40]). As such, Hypothesis 2 was supported.

A subsequent regression analysis tested whether background setting had a significant direct effect (c' path) on final interview scores, while accounting for the mediating variable, initial impression scores. An omnibus test of the direct effects indicated that after accounting for initial impression scores, background setting accounted for a very small and non-significant (.05%) of variance in final interview scores. Compared to the personal background condition, the professional ($b = -0.122$, $SE = 0.132$, $p = .355$, 95% CI[-.38, .14]), and the blurred ($b = -0.058$, $SE = 0.131$, $p = .662$, 95% CI[-.32, .20]) background conditions did not result in significantly higher final interview scores. Therefore, Hypothesis 3 was not supported, as background setting ultimately had no significant direct effects on final interview scores.

Examination of the 'ab' path indicated that there were no significant indirect effects in the model. All of the bootstrapped confidence intervals for the relative conditional indirect effects of background setting on final interview scores included the zero value, thus indicating a lack of indirect effects in the model. As such, Hypothesis 4 was not supported, as initial impressions did not mediate the relationship between background setting and interview performance scores.

The overall model (c path) accounted for 78.9% of the variance in final interview ratings, and the model was significant, $R^2 = .789$, $F(7, 392) = 209.75$, $p < .001$. That said, it must

be stressed that, in this case, the vast majority of the total variance was not associated with background setting, but rather with response quality.

Moderators

NFCC did not significantly moderate the relationship between initial impressions and final interview scores. NFCC did not have a significant main effect on final interview scores ($b = 0.059$, $SE = 0.067$, $p = .382$, bootstrapped 95% CI[-.07, .19]), and there was no significant interaction between initial impression scores and NFCC ($b = 0.063$, $SE = 0.066$, $p = .339$, bootstrapped 95% CI[-.07, .19]). As such, Hypothesis 5 was not supported.

Response quality showed a significant main effect in the overall model ($b = 4.04$, $SE = 0.108$, $p < .001$, bootstrapped 95% CI[3.83, 4.25]). There were no significant interactions between initial impression scores and response quality ($b = -0.060$, $SE = 0.106$, $p = .575$, bootstrapped 95% CI[-.27, .15]). Therefore, Hypothesis 6 was not supported, as applicant response quality did not moderate the relationship between initial impressions and final interview scores.

Feeling of Rightness

To answer Research Question 1, IBM SPSS Statistics PROCESSv4.0 Macro (Model 1; Hayes, 2017) was used to test whether high/low feelings of rightness were associated with a stronger/weaker relationship between initial impressions and final interview scores. In this case, initial impression scores represented the independent variable, feelings of rightness represented the moderator, and final interview scores represented the dependent variable.

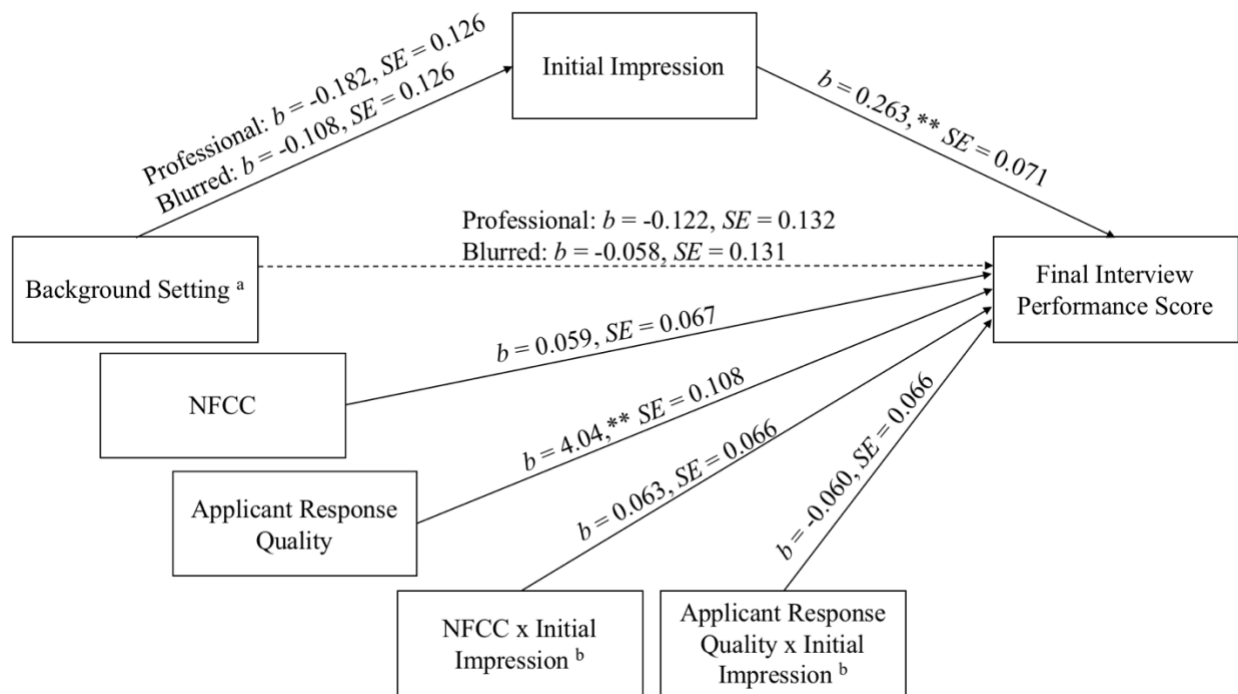
The overall model was significant ($R^2 = .059$, $F[3, 396] = 8.30$, $p < .001$), and accounted for 5.9% of the variance in final interview scores. Both initial impression ($b = 0.363$, $SE = 0.112$,

$p = .001$) and feeling of rightness ($b = -0.836, SE = 0.225, p < .001$) ratings had significant main effects on final interview scores.

Next, I evaluated the interaction between initial impression scores and feeling of rightness ratings. There was a non-significant overall incremental effect of the interaction, $b = -0.264, SE = 0.268, p = .324$. Therefore, higher feelings of rightness were not associated with a stronger relationship between initial impressions and final interview scores. Likewise, low feelings of rightness were not associated with weaker relationships between initial impressions and final interview scores.

Figure 2

Conceptual Model with Unstandardized Regression Weights and SEs



Note. The dotted line represents the direct effect (c' path) of background setting on final interview scores. NFCC = need for cognitive closure. $**p < .001$.

^a Refers to the background setting conditions, using the personal background condition as a reference group.

^b The 'x' indicates an interaction between two variables.

Exploratory Analyses

VCI interviewer experience as a covariate. I conducted an exploratory analysis in order to control for whether or not participants had previously conducted a video conference interview (VCI). I initially considered several other potentially relevant control variables relating to interview experience, including whether participants had previously participated in an AVI, rated an AVI, or conducted face-to-face (F2F) interviews. I ultimately chose VCI interviewer experience as the covariate, as it appeared to be the variable that was most representative of the various facets of interview experience, since it was the only variable of the four under consideration that was significantly correlated with all three of the other interview experience variables (AVI applicant experience: $r = .146, p < .01$; AVI interviewer experience: $r = .250, p < .01$; F2F interviewer experience: $r = -.106, p < .05$). VCI interviewer experience also showed stronger correlations with the outcome variables used in the model compared to the other three interview experience variables. VCI interviewer experience shared significant and moderately-high correlations with initial impression scores ($r = .165, p < .01$), and moderately-low correlations with final interview scores ($r = .092$), which suggests that VCI interviewer experience had potential to influence the overall outcomes of the study.

On a theoretical basis, VCI rating experience could also be considered an appropriate covariate based on the fact that those more experienced with VCIs may have been more habituated to using a computer modality to conduct interviews than participants without this particular type of experience. VCIs share some technological similarities to AVIs, in that the

interviews are conducted online using a web-cam, and the interviewer is able to view the candidates' backgrounds on the screen. Participants with VCI interviewer experience may have therefore been more comfortable with participating in more technological interviews than their counterparts who had only experienced F2F interviews. This experience and comfort assessing computer-based interviews could potentially have influenced the extent to which they were able to notice and incorporate the applicants' background into their ratings, as they may have spent less time and effort getting comfortable with the interview modality itself. VCI rating experience was therefore controlled for in order to rule out any potential confounding effects related to this variable.

VCI interviewer experience was included as a covariate within the same PROCESS model (Model 16) used in the main study analysis. The analysis was bootstrapped using 5000 samples, and continuous variables were again mean-centered. Controlling for VCI interviewer experience generally did not influence the overall results of the study. The 'a' path model became significant ($R^2 = .067$, $F[13, 374] = 2.06$, $p = .016$), but the professional ($b = -0.173$, $SE = 0.125$, $p = .168$, bootstrapped 95% CI[-.42, .07]) and the blurred ($b = -0.104$, $SE = .125$, $p = .406$, bootstrapped 95% CI[-.35, .14]) background conditions did not have significant main effects, in comparison with the personal background condition. Initial impressions ceased to be associated with final interview scores in the 'b' path ($b = 0.120$, $SE = 0.176$, $p = .494$, bootstrapped 95% CI[-.22, .47]). A test of the direct effects (c' path) indicated that the professional ($b = -0.138$, $SE = 0.134$, $p = .303$, bootstrapped 95% CI[-.40, .13]) and the blurred ($b = -0.085$, $SE = 0.134$, $p = .527$, bootstrapped 95% CI[-.35, .18]) background conditions again did not significantly influence final interview scores, in comparison with the personal background condition. All of the bootstrapped confidence intervals for the relative conditional

indirect effects of background setting on final interview scores ('ab' path) once again included the zero value, thus indicating a lack of indirect effects. The total effect model (c path) accounted for 78.8% of the variance in the final interview ratings, thus accounting for 0.1% less variance than that found in the original study analysis, $R^2 = .788$, $F(8, 384) = 178.87$, $p < .001$. Response quality ($b = -4.03$, $SE = 0.110$, $p < .001$, bootstrapped 95% CI[-4.25, -3.82]) continued to have a significant main effect, however, NFCC again did not ($b = 0.068$, $SE = .068$, $p = .318$, bootstrapped 95% CI[-.07, .20]). Neither response quality ($b = 0.072$, $SE = 0.108$, $p = .507$, bootstrapped 95% CI[-.14, .28]) nor NFCC ($b = -0.069$, $SE = 0.066$, $p = .296$, bootstrapped 95% CI[-.06, .200]) significantly interacted with initial impressions.

I also included VCI interviewer experience as a covariate in the moderation model used to assess the potential moderating effects of feeling of rightness, as it seemed plausible that familiarity with computer-based interviews may also have influenced participants' confidence in their ratings. VCI interview experience was the interview experience variable that was most strongly associated with feeling of rightness ratings, $r = -.095$. VCI interview experience was entered as a covariate into the same model that was previously used to assess Research Question 1 (Model 1).

The results of the exploratory analysis indicated that including VCI interviewer experience did not change the outcome of the moderation model. The overall model was significant ($R^2 = .059$, $F[4, 388] = 6.05$, $p < .001$), and both initial impression ($b = 0.339$, $SE = 0.115$, $p = .004$) and feeling of rightness ($b = -0.795$, $SE = 0.232$, $p < .001$) ratings had significant main effects on final interview scores. There also continued to be a non-significant overall incremental effect of the interaction, $b = -0.265$, $SE = 0.270$, $p = .327$.

Controlling for remote work experience and access to a private and dedicated home workspace (PDHW). I also considered several other potentially relevant control variables relating to remote work experience, including whether participants were currently working remotely, had previously worked remotely, whether participants were working remotely due to COVID, and whether or not participants had a private and dedicated home workspace (PDHW). These variables were considered as possible control variables based on the possibility that those working remotely may reflect on their own experiences and be more lenient when rating applicants adapting to a computer-mediated modality and a lack of access to a more professional workspace.

I initially chose PDHW as the covariate, as it appeared to be the variable that was most representative of the various facets of remote work experience. This decision was based on the fact that PDHW was the only variable of the four under consideration that was significantly correlated with all three of the other remote work variables (current remote work: $r = .257, p < .01$; previous remote work experience: $r = .207, p < .01$; remote work due to COVID: $r = .169, p < .01$). I subsequently assessed the relationships between PDHW and the original study variables, but found that PDHW shared almost no relationship with them (i.e. initial impression scores: $r = .029$, final interview scores: $r = -.008$). The remaining remote work variables similarly showed almost no association with any of the original study variables. Despite the reasoning to control for remote work experience and PDHW, the lack of association between these variables and the original study variables suggested that there was little reason to conduct further covariate analyses (Bernerth & Aguinis, 2016).

Discussion

Despite the rapidly growing popularity of AVIs, little research has previously examined the potentially biasing effects that visible backgrounds in AVI recordings may have on final interview outcomes. This study addressed this shortfall, and specifically contributes to the literature by assessing how AVI background settings may influence initial impressions and final interview scores, using the dual-process theory as a theoretical framework. Overall, the results indicated that final interview scores were influenced by initial impressions and the quality of applicant responses. In addition, the study showed that recording an AVI from a personal space, like a bedroom, was perceived as being less professional than recording an interview in a home office environment. Despite this difference, however, background setting was not associated with initial impression scores or final interview scores, and initial impressions did not significantly mediate the model. There were no significant interactions between initial impressions and NFCC, or between initial impressions and response quality. Feelings of rightness also did not significantly influence the relationship between initial impressions and final interview scores. Despite the fact that the results only supported one of the study's hypotheses, the results of the study nevertheless point to several interesting and relevant conclusions for applicants, organizations and researchers involved in AVIs.

Theoretical Implications

Influence of background setting. This study's primary focus was to determine whether or not the background setting influenced the initial impression and final interview scores awarded to the AVI applicant. Although background setting did not have a notable effect on either initial impressions or final interview scores, the findings nevertheless suggest a number of potential theoretical explanations, which merit further investigation in future research.

First, the background settings' lack of effect on initial impressions and final interview scores could be attributed to the current study's operationalization of Type 1 and 2 processes in initial impression formation. Previous research showed that individuals automatically and immediately process visual stimuli to form initial impressions (Thorpe et al., 1996), and it is therefore possible that the current study allotted more time than was necessary to form initial impressions. The introductory videos used in the present study were approximately 13 seconds in length, which should have allowed for immediate processing of background visual cues, but may also have allowed participants sufficient time to engage in Type 2 processing. If participants had the opportunity to engage in more Type 2 processing, they may have re-examined their initial impressions more carefully, and regulated their original judgements (Derous et al., 2016). In other words, participants may have had the opportunity to consciously reflect on the fact that the background setting had no bearing on the applicant's suitability for the position, resulting in the background settings' ultimate lack of effect on initial impression and final interview ratings. Therefore, in line with Evans' (2012) previous call to resolve certain theoretical challenges associated with the dual-process theory framework, future research may benefit from more closely examining the amount of time that it takes for individuals to engage in Type 2 processing, and the factors that may influence conscious decision-making processes.

Another potential theoretical explanation for the lack of background setting effects may be described in terms of the load theory of selective attention and cognitive control. Load theory suggests that an individual's ability to process extraneous information is dependent on the resources that are available to interpret it. Two selective attention mechanisms, namely perceptual selection and cognitive control mechanisms, regulate whether or not an individual processes distractor information (Lavie et al., 2004). Perceptual selection is thought to occur in

cases of high perceptual load, when there is insufficient capacity to process irrelevant distractor information. In such cases, irrelevant details simply go unnoticed as the individual processes other incoming information. Cognitive control involves higher cognitive functions (such as working memory) to continuously maintain ongoing processing priorities. When there is low load on cognitive functions, individuals are thought to be able to maintain their focus on processing priorities.

Therefore, in the context of the present study, it is possible that the short introductory videos contributed to a state of high perceptual and low cognitive load, thus preventing background setting from influencing initial impression formation. For instance, simultaneously adapting to an AVI format and listening to the applicants' responses may have created a state of high perceptual load, which may have prevented raters from noticing the background content. This supposition is further supported by the fact that, in response to an exploratory manipulation check item that asked participants to "please indicate what you believe the present study is focused on," one participant responded "...I am not sure what the video background questions were all about. I did not notice the background AT ALL. I focused on the candidate." Furthermore, there were more pronounced differences in professionalism ratings across the professional and personal conditions in the pilot study still photos than were found using the main study videos. This difference suggests that the participants may have noticed the background in the still photos more than in the videos, which could also be indicative of the greater perceptual load caused by verbal responses in the video conditions. As such, it is plausible that the background setting manipulation simply went unnoticed in the context of the present study due to participants' focusing on the verbal responses. Even though the hypotheses

were not supported, the fact that background setting did not have a notable effect is in itself promising, however, as it suggests that less bias occurred when raters made their judgements.

In contrast, the background settings' lack of influential effect could also relate to the degree of cognitive control experienced by the study participants. Passively evaluating a short AVI introductory video may have placed a low load on cognitive functions, as the raters simply had to watch and assess the recorded videos rather than simultaneously conduct and rate the interviews (as they would in F2F contexts). As such, even if the raters happened to notice the background setting in the introductory videos, the low cognitive load may have allowed them to maintain their focus on the interview responses (the processing priority), thus effectively ignoring the background setting content. Therefore, the introductory videos in the present design may have inadvertently caused raters to experience either a state of high perception or low cognitive load (or both), thus preventing the background setting manipulation to be fully noticed or integrated into initial impression ratings. Further examination of the perceptual and cognitive load mechanisms inherent to load theory in an AVI context may therefore point to a fruitful area of future exploration in AVI research. Regardless, however, the results suggest that raters were able to focus on the most important and relevant information needed to distinguish strong from weak candidates, and therefore show promise for selection processes relying on the AVI modality.

Background setting may also have failed to achieve the anticipated initial impression and final interview outcomes due to changing normative behaviours, perceptions of control and the degree of visibility associated with AVI background content. Although recording an AVI from a personal space, such as a bedroom, may have at one time been viewed as a breach of normative behaviour, it is possible that modern-day COVID-19 circumstances may have altered perceptions

as to what constitutes an acceptable space to complete an AVI in. Since a large portion of the workforce suddenly found themselves working from home when the pandemic began, a substantial number of people were forced to adapt their personal living spaces into areas in which they could work remotely (Kniffin et al., 2020). Many workers simply did not have access to a private, designated professional workspace in their homes, and as such, a number of anecdotal sources attest to people resorting to working from bedrooms, living rooms and even closets during periods of COVID-19 lock-down (i.e. Kniffin et al., 2020; Vassel, 2020). Similarly, although concealing one's background may at one time may have suggested that the individual was hiding something (Maddocks, 2021), the use of blurred backgrounds has also become more common since video conferencing platforms started integrating blur features into their platforms during the pandemic (i.e. Raney, 2021). The sudden and global change that came with the pandemic may therefore have shifted previous perceptions of normative behaviour, in that participating in remote conferencing from a personal living space or using blurred backgrounds ceased to be viewed as a negative deviation from the norm, and simply became the new normal.

As such, the background setting conditions may not have had the anticipated effects, simply because none were viewed as being a deviation from the new norm. Without the perception of there being an abnormality, a personal background may not have triggered the negative affective reactions (i.e. uncertainty or discomfort, etc.) inherent to immediate Type 1 responses, which are thought to influence the raters' initial impression formation (Derous et al., 2016). Instead, in the absence of affective reactions, the raters may have engaged in more Type 2 conscious reasoning and thoughtfully considered the AVI applicant's suitability for the target position. Type 2 conscious reasoning would likely lead raters to conclude that an applicant's background bears little relevance to job suitability, thereby counteracting the potential biasing

effects of background setting. Further examination of normative behaviours in light of the pandemic are therefore recommended in future AVI research.

Relatedly, the pandemic may also have changed perceptions of applicant control and background visibility, which are thought to influence the degree of negativity associated with stigmatizing features. As such, the extent to which the recording location was perceived to be within the applicant's control and the visibility of the content may have reduced the stigmatizing effect of the background setting, even if personal spaces continued to be viewed as a deviation from the norm. Based on the fact that so many workers worldwide were suddenly required to work from home at the onset of the pandemic, many may now view a personal background as being outside of an applicant's control, which may have reduced the negativity typically associated with recording an AVI in such a location.

The perception of there being a lack of control may even have been more influential than the degree of visibility in terms of its effect on the negativity associated with stigmatizing features. For instance, in the current study, manipulating the visibility of the background did not influence rater judgements, as the blurred background condition did not have any more of an effect on initial impressions and final interview scores than the visible professional and personal background conditions did. Therefore, in addition to exploring the pandemic's effects on normative behaviours, further examination of how COVID-19 circumstances may have influenced the interplay of background controllability and visibility may therefore also be warranted in future AVI research.

Factors that may influence negative initial impressions. Future research may also benefit from considering other background feature factors that may be perceived as a greater deviation from the norm or more controllable than background setting. Visible indicators of

protected group status may represent one such influential background feature. The background manipulations that were included in the present study were neutral, in the sense that they were devoid of objects that indicated protected group status or other personal information. It is therefore possible that background setting in itself is simply not as polarizing or biasing a factor as viewing evidence of family status, religious affiliation, sexual orientation or low socio-economic status. Future research should therefore continue to examine the potential biasing effects that visual indicators of protected group status may have on final AVI outcomes.

Similarly, raters may draw more meaningful inferences from the cleanliness of the background space compared to the type of background setting. The background conditions in the present study were consistently tidy and clean, and therefore may have been perceived positively by raters, regardless of the type of setting associated with the recorded space. A less tidy or clean space, for instance, may indicate that the applicant is not a conscientious individual (Gosling et al., 2002; Lee & Ashton, 2009), which may have a greater impact on perceptions of the applicant's job suitability. Inferences regarding an applicant's level of conscientiousness may therefore be viewed as more salient and a greater deviation from the norm than inferences drawn from background settings, thus triggering more negative Type 1 affective responses. As such, the importance of and inferences drawn from background cleanliness or tidiness may also warrant further investigation.

In addition to protected status or cleanliness, applicant appearance may also have eclipsed the effect that background setting had on initial impressions and final interview scores. Prior research previously noted that applicant appearance had an influential effect on interview ratings, in that physical attractiveness and/or style detracted from final interview judgements (Suen et al., 2019; Torres & Gregory, 2018). For example, Suen et al. (2019) found that both

applicant appearance and initial impressions were highly correlated with interview scores, which suggests that applicant appearance may have a more important influential effect than background setting in an AVI context. The AVI applicant in the present study was well-groomed and wore professional attire to record the AVI responses, and this fact was also noted by several participants in the study when they were asked what they believed the intent of the study was. These comments suggest that the AVI applicant's appearance and attire were deemed noticeable enough by participants to warrant mentioning them. It is therefore possible that the AVI applicant's overall professional appearance in the present study dispelled the effects of the background conditions. Examining the role of applicant appearance in relation to background setting may therefore also represent a valuable future research avenue. In addition, researchers could examine whether casual attire and appearance magnify the biasing effect of a personal background (and vice versa), and whether pairing professional attire with a personal background might neutralize such effects. In summary, there are a number of potential biasing background factors that should be more closely examined in future research to better understand their potentially negative influential effects on AVI interview bias.

Influence of initial impressions. Similar to findings pertaining to traditional F2F interviews, the results suggest that initial impressions influence final interview outcomes. The effect of initial impressions on final interview scores lends support to Tversky & Kahneman's (1974) description of anchoring bias, in that individuals tend to avoid significant deviations from their initial assessments. These findings also support the initial impression component of Derous et al.'s (2016) model, which placed heavy emphasis on the anchoring role of initial impressions in interview settings.

Consistent with anchoring bias, higher initial impressions were associated with higher final interview scores, thus reinforcing previous findings that initial impressions may represent a form of bias in AVI selection contexts (Suen et al., 2019). Even though the background setting conditions did not in themselves introduce bias to the selection process in the current study, it is clear that there is still opportunity for biased initial impressions to have an effect on final interview scores in an AVI context. This information is particularly relevant, given that AVIs' enhanced interview structure and lack of interviewer rapport-building, prompting and probing have been touted for their potential to reduce the amount of bias introduced into selection processes (HireVue, 2019; Lukacik et al., 2022). The expectation that AVIs are less biased, paired with the lack of background setting effects found in the current study, run the risk of researchers and practitioners discounting the potential adverse effects related to other biasing visual background features. These findings therefore further reinforce the need for future research to examine other potentially biasing visual background features, which could have a greater influence on initial impressions, and subsequently, final interview scores.

Influence of feeling of rightness. Despite the fact that initial impressions appear to anchor final interview scores, the precise mechanism associated with this anchoring effect remains unclear, based on the outcomes of the study. Although Derous et al.'s (2016) model suggests that feelings of rightness represent the anchoring process that relates initial impression ratings to final interview scores, the results did not support this supposition. Contrary to expectations, feeling of rightness shared almost no relationship with initial impression scores, and feelings of rightness did not moderate the relationship between initial impressions and final interviews scores.

That said, there is some evidence that suggests that the interaction may simply have gone undetected by the present study, which suggests that additional research should be conducted before dismissing the potential anchoring role that feelings of rightness play in AVI rating decisions. First, it should be noted that although feelings of rightness did not moderate or share a notable relationship with initial impressions, feeling of rightness scores had a significant main effect on final interview scores, and shared moderately-high significant negative relationships with response-level ratings, final interview scores and applicant response quality. In other words, contrary to Derous et al.'s (2016) predictions, participants who evaluated the applicants negatively appeared to be more confident in their evaluations than those who evaluated applicants positively. Despite the lack of interactions in the present study, another study found that feeling of rightness ratings showed moderating effects in a traditional F2F interview context, using similar measures to assess feelings of rightness levels (Buijsrogge et al., 2016). This finding suggests that perhaps differences exist between F2F and AVI contexts that influence the moderating role of feelings of rightness. The fact that Suen et al. (2019) found a smaller initial impression anchoring effect in AVIs compared to synchronous video interviews further suggests that an interaction between initial impressions and feelings of rightness may simply be more challenging to detect in an AVI context. Future studies should therefore explore how differences in interview contexts may affect the moderating role that feelings of rightness may fulfill.

Influence of applicant response quality. The results of the study also support previous research findings that point to the importance of applicants' response quality in interview settings. Although the quality of responses is known to play an important role in traditional and simulated in-person final interview scores (Catano et al., 2022; Roulin & Bhatnagar, 2021), the present study demonstrates that response quality also plays a notable part in determining

interview outcomes in an AVI context. This finding could have important theoretical implications for researchers comparing the validity of different interview modalities, in that AVI outcomes still appear to primarily reflect the quality of applicant responses provided in the recorded videos. The importance of response quality may indirectly highlight the fact that applicants' skills, experience, knowledge, qualifications and job fit are more predictive of final interview outcomes than other potentially biasing factors in an AVI context, such as background setting. This outcome was encouraging and lends credibility to the AVI modality, as it suggests that regardless of other potential biasing effects included in the study, information that is pertinent to an individual's suitability for a given position (Catano et al., 2022) carried the most weight in final interview decisions.

These findings may also have important implications for research examining AVI candidate reactions. Applicants concerned about the lack of interpersonal interaction and opportunity to impression manage (Basch et al., 2020) may be reassured by the fact that their response quality nevertheless represents an important factor in determining final interview outcomes. As such, explaining to candidates that response quality ultimately carries more weight than certain other factors in rendering final interview decisions may improve applicants' generally negative outlook towards AVIs (as previously observed by Basch & Melchers, 2019; Langer et al., 2021).

Although response quality had a direct effect on final interview scores, it did not significantly interact with initial impressions. This finding was surprising, given that Roulin & Bhatnagar (2021) found that the anchoring effect of initial impressions was weaker for high quality responses compared to average quality responses. Since the present study used more polarized levels of response quality (i.e. high- and low-quality responses) than those

implemented by Roulin & Bhatnagar (2021), a more pronounced moderating effect was initially expected, rather than the lack of significant effect that was observed.

An explanation for this discrepancy may relate to Roulin & Bhatnagar's (2021) theoretical interpretation of the response quality's moderating effect. Roulin & Bhatnagar (2021) proposed that the moderating effect found in their study was likely attributable to raters engaging in more Type 2 processes when rating high quality responses, and a greater reliance on initial impressions and Type 1 processes when evaluating lower quality responses. By extending this logic to the lack of moderating effect in the present study, it is possible that the current study's participants more consistently engaged in Type 2 processes across the two response quality conditions, thus nullifying the moderating effect of response quality. For instance, the more polarized responses may have been so easy to identify that participants only focused on the quality of the responses, thus ignoring other extraneous information in the AVI. In other words, the high and low quality response conditions may have eased Type 2 processing, which may have subsequently nullified any potential biasing effects.

Influence of NFCC. NFCC also failed to moderate the relationship between initial impressions and final interview scores. This finding was not overly surprising, given the overall lack of stigmatizing effect that background setting had on initial impressions and final interview scores. Essentially, there was likely little opportunity for NFCC to moderate raters' responses to stigma in the present study, given that the stigmatizing feature, background setting, ultimately had no effect on the other variables examined within the overall experiment. Given that there were no apparent stigmatizing features to fixate on during initial impression formation, raters likely focused more on the verbal responses provided by the AVI applicant. Raters therefore likely engaged in more Type 2 conscious information processing across all of the background

setting conditions. In doing so, NFCC would have had little opportunity to differentially influence the anchoring effects of initial impressions.

As such, the fact that NFCC failed to directly influence final interview scores or interact with initial impressions in the present study should not discourage researchers from considering NFCC in future studies. Previous studies examining traditional F2F interviews (i.e. Alison et al., 2008; Buijsrogge et al., 2021), for instance, found notable effects of NFCC in relation to stigmatizing features using the same measure as that used in the present study (Buijsrogge et al., 2021). As such, there nevertheless appears to be value in further examining NFCC in AVI contexts, perhaps with some adjustments in terms of the stigmatizing feature being examined. Future research focusing on stigma in AVI background environments should thus consider including NFCC as a potential moderator of initial impressions, but should incorporate alternative stigmatizing features in their study design. Some such features could include those previously mentioned, such as items indicating protected group status, untidy background spaces or unprofessional attire or appearances, for instance.

Practical Implications

The study results also point to a number of pertinent practical implications for AVI applicants and organizations considering incorporating AVIs into their selection processes. First, both applicants and organizations should note that background setting alone does not appear to dramatically influence either rater initial impressions or final interview judgements. This finding is of particular interest, given that many AVI platforms tout the fact that applicants can record their AVIs from any location with internet access, thus offering increased flexibility and convenience for job seekers (HireVue, 2019; Modern Hire, 2020). The lack of influence that background setting exerted in the present study therefore suggests that from a background setting

standpoint, applicants can in fact record AVIs in a bedroom or home office setting, even if personal spaces have traditionally been considered unprofessional. Likewise, using a blurred background does not appear to place candidates at a greater disadvantage than when a visible home office or bedroom setting is used. As such, AVIs likely offer a degree of flexibility in terms of where interviews are completed without adversely affecting applicants' interview outcomes. This flexibility may benefit job candidates coordinating work and personal schedules, as well as widen the organization's applicant pool to candidates living outside of the geographical location (Catano et al., 2022; HireVue, 2019).

Despite this promising finding, however, it should be noted that this study only examined the setting of the background space, and therefore interpretation of the findings can only be applied to the type of background setting in which applicants choose to complete their AVIs. As such, applicants are cautioned to consider other possible background features apart from setting that may inadvertently affect their interview outcomes prior to recording their AVI videos. For instance, applicants should consider how the tidiness, cleanliness, lighting, background noise and personal items might signal potentially negative or sensitive information to the rater (Cummins, 2021; Myrick, 2021; Vasel, 2021). When using a blurred background, applicants should also consider the proximity of any undesirable features to the camera, which may become visible by virtue of being close to the screen or when the candidate moves around (Maddocks, 2021). If in doubt and whenever possible, applicants are advised to remove or alleviate any potentially stigmatizing features, or consider a more appropriate space to conduct their AVI in.

Although background setting did not influence initial impressions, the study results nevertheless suggest that the anchoring effect observed in traditional interviews also appears to persist in AVIs. The present study found that initial impressions were predictive of final

interview scores, which suggests that applicants should bear in mind the manner in which they present themselves, as they would in a traditional interview. As previously mentioned, little is currently known about possible interactions between visual background features, and therefore the effects of background setting may become magnified or understated when combined with other factors, such as professional attire and appearance. Until more is known about other factors influencing initial impressions, applicants should therefore consider erring on the side of caution, and strive to maximize the likelihood of obtaining positive initial impressions by presenting themselves as professionally as possible in an AVI.

To complement existing AVI best practices, the study findings also provide some preliminary guidance in terms of how applicants may select their AVI backgrounds. Although it should again be emphasized that background setting did not ultimately have an effect on final interview ratings, it is worth noting that there were also no differences in terms of how professional the raters perceived the personal and blurred backgrounds to be. These findings suggest that if applicants do not have access to a professional space, blurring their backgrounds to conceal a personal background likely will not improve nor worsen rater perceptions of the background setting. Therefore, the background manipulation included in the present study nevertheless provides some useful guidance for AVI candidates who are debating about whether or not to blur their backgrounds to improve their ratings.

The importance of initial impressions in an AVI context are also relevant for organizations planning to integrate AVIs into their selection processes. AVI platforms and some researchers have proposed that AVIs reduce biased initial impressions by reducing the amount of ancillary information normally shared during traditional interviews (Basch & Melchers, 2019; Gorman et al., 2018; HireVue, 2019; Lukacik et al., 2022). The results of the present study

indicate, however, that initial impressions continue to have an important influence on final interview outcomes in AVI contexts. As such, organizations should strive to minimize the amount of ancillary information shared by the applicant at the forefront of the interview, particularly with regards to the information that candidates are asked to share during introductory statements. To limit the amount and type of non-job related information shared by applicants (Levashina et al., 2014), whenever possible, organizations should avoid including introductory statements in their AVI selection processes. If organizations allow applicants to record introductory statements, such recordings should not be provided to raters. These measures may minimize the effect that introductory statements have on rater initial impression formation.

To further maximize selection process fairness and reduce the anchoring effects of initial impressions, organizations could also consider standardizing candidate backgrounds and utilizing structured interview rating systems. Although blurred backgrounds did not have a meaningful effect as a manipulation of setting in the present study, the results of the pilot study suggest that the blur function might address other concerns related to background content and visibility. The pilot study results indicated that raters are not able to identify objects when the blur function was used, and as such, requiring that applicants blur their backgrounds may prevent raters from viewing items of a potentially sensitive or protected group nature. Such measures may therefore reduce the influence of rater biases that could lead to negative initial impression formation, and ultimately contribute to unfair selection processes. In addition, asking applicants to blur their backgrounds may also alleviate candidate concerns regarding invasions of their privacy and improve overall applicant reactions towards AVIs, as their surroundings will no longer be visible to raters. It should be noted, however, that despite the potential practical benefits that blurred backgrounds may offer, many AVI platforms do not yet provide a background blur function, and

additional research is needed to confirm that blurring backgrounds will yield the proposed advantages.

Limitations

Although this study offers a number of interesting implications for researchers, applicants and organizations, it is not without some limitations. First, the study utilized a simulated selection context, and therefore the results may not be generalizable to real selection contexts (Blacksmith et al., 2016). In addition, the high and low quality response scripts were quite polarized, in that the high quality responses represented extremely strong responses, and the low quality response scripts reflected extremely poor responses. Such polarized responses may not realistically be representative of typical AVI responses, which would likely include more average quality responses than those included in the study. Therefore, future research would likely benefit from further examining AVI biases using a real selection context and more varied candidate responses to gain a better understanding as to how well the present study's findings generalize to actual AVI selection process contexts.

Second, the AVI videos recorded for the study were recorded individually, rather than leveraging studio and software to manipulate the background. Although I initially considered utilizing a recording studio and professional software to manipulate the background conditions, the available options were ultimately deemed unsuitable, as the videos did not appear to be realistic. As a result, the mock AVI applicant recorded each video twice in an effort to minimize any differences in eye contact, verbal mannerisms, body movement, lighting or audio features across the videos. Despite these efforts, however, it is possible that minor differences may have gone unnoticed, and that differences between the videos introduced some unknown confounding factors into the study results. Future studies with more advanced resources and video-editing

skill levels at their disposal should therefore ideally consider recording the video responses in a studio setting, and then manipulating the background in order to further minimize any differences between the conditions.

Finally, although participation in this study required some interview experience, the participants reported a wide range in experience levels. Most had never participated in an AVI before, and experience conducting F2F interviews ranged from having conducted only a few interviews to several hundred in the past. This wide range in experience levels may have influenced raters' degree of familiarity, comfort and relative ease in rating the AVI responses, which may have influenced the overall results and the extent to which the backgrounds were noticed. Additionally, given the general lack of experience with AVIs, the participants in the present study may have benefited from a greater opportunity to become accustomed to this interview modality before being asked to rate the candidate's responses. Future research may therefore benefit from selecting a participant pool with a more homogenous level of human resources or interview experience, and should consider providing an opportunity to view the AVI platform and practice evaluating the video responses.

Conclusion

The rapidly increasing prevalence of AVIs in selection processes worldwide highlights the need to better understand the possible ramifications of this interview modality. As a number of researchers (i.e. Lukacik et al., 2022) have identified, technological attributes inherent to AVIs, such as the introduction of visible environmental features in the background of recordings, have the potential to disproportionately influence interview outcomes. Yet, despite the risk of negative consequences that may exist for both organizations and applicants, very little is known in terms of the possible biasing effects that visual background cues may introduce to the

selection process. Using components of the dual-process framework proposed by Derous et al. (2016), this study therefore examined the effects of background setting on initial impression and final interview outcomes.

This study contributes to the literature by being among the first to test some of the elements included in the dual-process theory framework outlined by Derous et al. (2016), specifically in the context of bias introduced by background cues in AVIs. In doing so, this study supports previous research findings pointing to the importance of response quality and the anchoring effects of initial impressions, and suggests that the type of background setting does not in itself adversely influence initial impression or final interview outcomes. These findings point to future promising avenues of research, which may allow for a deeper understanding of the theoretical mechanisms involved in interview bias, and identify other AVI background features that may have influential effects on interview outcomes. As such, the present study takes initial steps to better understand the underlying processes involved in AVI judgement formation and the extent to which certain AVI technological attributes influence interview outcomes.

The results of this study also provide some practical guidance to both organizations and applicants exploring AVI processes. This study may help inform candidates concerned about the environment in which they opt to complete their AVI recordings, and the results complement existing organizational guidance in terms of how to minimize biased initial impressions on the part of interview raters and promote fair AVI selection processes. In sum, this research provides some additional information to applicants and organizations exploring AVI best practices, and research seeking to better understand the potential biasing effects inherent to AVI-based selection processes.

Although there remains much to be learned in terms of bias in AVI selection contexts, research is steadily uncovering the mechanisms and processes involved in bias formation related to the AVI modality. Overall, this study contributes to the growing literature on AVIs, and points to the need for additional research in order to better understand, and ultimately alleviate, potential sources of AVI bias. In doing so, this study advances the overall knowledge of AVI attribute effects in order to promote fair, technologically-mediated interviews in the future.

References

- Alison, L., Kebbell, M., & Leung, J. (2008). A facet analysis of police officers' self-reported use of suspect-interviewing strategies and their discomfort with ambiguity. *Applied Cognitive Psychology, 22*(8), 1072–1087. <https://doi.org/10.1002/acp.1408>
- Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology, 91*(4), 652–661. <https://doi.org/10.1037/0022-3514.91.4.652>
- Aronow, P. M., Baron, J., & Pinson, L. (2019). A note on dropping experimental subjects who fail a manipulation check. *Political Analysis, 27*, 572–589. <https://doi.org/10.1017/pan.2019.5>
- 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. <https://doi.org/10.1037/a0016532>
- Basch, J., & Melchers, K. (2019). Fair and flexible?! Explanations can improve applicant reactions toward asynchronous video interviews. *Personnel Assessment and Decisions, 5*(3). <https://doi.org/10.25035/pad.2019.03.002>
- Basch, J. M., Melchers, K. G., Kegelmann, J., & Lieb, L. (2020). Smile for the camera! The role of social presence and impression management in perceptions of technology-mediated interviews. *Journal of Managerial Psychology, 35*(4), 285–299. <https://doi.org/10.1108/JMP-09-2018-0398>

- Bernerth, J. B., & Aguinis, H. (2016). A critical review and best-practice recommendations for control variable usage. *Personnel Psychology, 69*(1), 229–283.
<https://doi.org/10.1111/peps.12103>
- Blacksmith, N., Willford, J., & Behrend, T. (2016). Technology in the employment interview: A meta-analysis and future research agenda. *Personnel Assessment and Decisions, 2*(1).
<https://doi.org/10.25035/pad.2016.002>
- Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology, 100*(2), 431–449.
<https://doi.org/10.1037/a0038047>
- Buijsrogge, A., Deros, E., & Duyck, W. (2016). Often biased but rarely in doubt: How initial reactions to stigmatized applicants affect interviewer confidence. *Human Performance, 29*(4), 275–290. <https://doi.org/10.1080/08959285.2016.1165225>
- Buijsrogge, A., Duyck, W., & Deros, E. (2021). Initial impression formation during the job interview: Anchors that drive biased decision-making against stigmatized applicants. *European Journal of Work and Organizational Psychology, 30*(2), 305–318.
<https://doi.org/10.1080/1359432X.2020.1833980>
- Cable, D. M., & Judge, T. A. (1997). Interviewers' perceptions of person-organization fit and organizational selection decisions. *Journal of Applied Psychology, 82*(4), 546–561.
<https://doi.org/10.1037/0021-9010.82.4.546>
- Catano, V. M., Hackett, R. D., Wiesner, W. H., & Roulin, N. (2022). *Recruitment and selection in Canada (8th ed)*. Top Hat.
- Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology, 50*(3), 655–702.

- Crocker J., Major B., & Steele, C. (1998). Social stigma. In Fiske, S., Gilbert, D., & Lindzey, G. (Eds.), *Handbook of Social Psychology: Vol. 2.* (pp. 504–53). McGraw-Hill.
- Cummins, L. (2021). *How to level up your home office background.* Anyvoo.
<https://anyvoo.com/home-office-background/>
- 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. <https://doi.org/10.1016/j.hrmr.2015.09.006>
- Derous, E., Ryan, A. M., & Nguyen, H.-H. D. (2012). Multiple categorization in resume screening: Examining effects on hiring discrimination against Arab applicants in field and lab settings. *Journal of Organizational Behavior, 33*(4), 544–570.
- Derous, E., Ryan, A. M., & Serlie, A. W. (2015). Double jeopardy upon resumé screening: When Achmed is less employable than Aïsha. *Personnel Psychology, 68*(3), 659–696.
<https://doi.org/10.1111/peps.12078>
- Dipboye, R. L., Fromkin, H. L., & Wiback, K. (1975). Relative importance of applicant sex, attractiveness, and scholastic standing in evaluation of job applicant resumes. *Journal of Applied Psychology, 60*(1), 39–43. <https://doi.org/10.1037/h0076352>
- Eisenkraft, N. (2013). Accurate by way of aggregation. *Journal of Experimental Social Psychology, 49*(2), 277–279. <https://doi.org/10.1016/j.jesp.2012.11.005>
- Evans, J. St. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378–395. <https://doi.org/10.3758/BF03193858>
- Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*(1), 255–278.
<https://doi.org/10.1146/annurev.psych.59.103006.093629>

Evans, J. St. B. T. (2012). Questions and challenges for the new psychology of reasoning.

Thinking & Reasoning, 18(1), 5–31. <https://doi.org/10.1080/13546783.2011.637674>

Gioia, D. A., & Poole, P. P. (1984). Scripts in organizational behavior. *Academy of Management*.

The Academy of Management Review (Pre-1986), 9(3), 449-459.

Goffman, E. (2009). *Stigma: Notes on the management of spoiled identity* (2nd Touchstone ed.).

Simon & Schuster.

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), 129–146. <https://doi.org/10.1037/cpb0000102>

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), 379–398. <https://doi.org/10.1037/0022-3514.82.3.379>

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis*

(2nd Ed.). Guilford Press.

Heatherton, T. F. (2003). *The Social Psychology of Stigma*. Guilford Press.

Hebl, M. R., & Kleck, R. E. (2002). Acknowledging one's stigma in the interview setting:

Effective strategy or liability? *Journal of Applied Social Psychology*, 32(2), 223–249.

<https://doi.org/10.1111/j.1559-1816.2002.tb00214.x>

Hemamou, L., Felhi, G., Martin, J.-C., & Clavel, C. (2019). *Slices of attention in asynchronous*

video job interviews. 8th International Conference on Affective Computing and Intelligent

interaction (ACII). <http://arxiv.org/abs/1909.08845>

- Higgins, C. A., & Judge, T. A. (2004). The effect of applicant influence tactics on recruiter perceptions of fit and hiring recommendations: A field study. *Journal of Applied Psychology, 89*(4), 622–632. <https://doi.org/10.1037/0021-9010.89.4.622>
- HireVue. (2019, May 21). *HireVue Surpasses Ten Million Video Interviews Completed Worldwide*. <https://www.hirevue.com/press-release/hirevue-surpasses-ten-million-video-interviews-completed-worldwide>
- Houston, V., & Bull, R. (1994). Do people avoid sitting next to someone who is facially disfigured? *European Journal of Social Psychology, 24*(2), 279–284. <https://doi.org/10.1002/ejsp.2420240205>
- Huffcutt, A. I., Roth, P. L., Conway, J. M., Stone, N. J., Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology, 86*(5), 897–913. <https://doi.org/10.1037/0021-9010.86.5.897>
- Kitchen, J. P., Kerr, G., Schultz, D. E., McColl, R., & Pals, H. (2014). The elaboration likelihood model: Review, critique and research agenda. *European Journal of Marketing, 48*(11/12), 2033–2050. <https://doi.org/10.1108/EJM-12-2011-0776>
- Kniffin, K. M., Narayanan, J., Anseel, F., Antonakis, J., Ashford, S. P., Bakker, A. B., Bamberger, P., Bapuji, H., Bhave, D. P., Choi, V. K., Creary, S. J., Demerouti, E., Flynn, F. J., Gelfand, M. J., Greer, L. L., Johns, G., Keesbir, S., Klein, P. G., Lee, S. Y., ... van Vugt, M. (2021). COVID-19 and the workplace: Implications, issues, and insights for future research and action. *American Psychologist, 76*(1), 63-77. <https://doi.org/10.1037/amp0000716>

Kruglanski, A. W. (2021). *The Need for Closure Scale (NFCS)*.

<https://www.kruglanskiarie.com/need-for-closure>

Jones, E. E., Farina, A., Hastorf, A. H., Markus, H., Miller, D. T., & Scott, R. A. (1984). *Social stigma: The psychology of marked relationships*. Freeman.

Landers, R. N., & Marin, S. (2021). Theory and technology in organizational psychology: A review of technology integration paradigms and their effects on the validity of theory. *Annual Review of Organizational Psychology and Organizational Behavior*, 8(1), 235–258. <https://doi.org/10.1146/annurev-orgpsych-012420-060843>

Langer, M., Baum, K., König, C. J., Hähne, V., Oster, D., & Speith, T. (2021). Spare me the details: How the type of information about automated interviews influences applicant reactions. *International Journal of Selection and Assessment*, 29(2), 154–169.

<https://doi.org/10.1111/ijsa.12325>

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, 133(3), 339–354.

<https://doi.org/10.1037/0096-3445.133.3.339>

Lee, K., & Ashton, M. C. (2009). The HEXACO Personality Inventory - Revised: A measure of the six major dimensions of personality. <https://hexaco.org/>

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. <https://doi.org/10.1111/peps.12052>

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). <https://doi.org/10.1016/j.hrmr.2020.100789>

- Madera, J. M. (2016). Facial stigmas in dyadic selection interviews: Affective and behavioral reactions toward a stigmatized applicant. *Journal of Hospitality & Tourism Research*, 40(4), 456–475. <https://doi.org/10.1177/1096348013503996>
- Maddocks, S. (2021). *Natural, virtual, or blurred backgrounds*. MSB Executive Blog. <https://www.msbexecutive.com/natural-virtual-or-blurred-backgrounds/>
- Mathis, R. L., Jackson, J. H., Valentine, S. R. (2014). *Human Resource Management (14th ed.)*. Cengage Learning.
- 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. <https://doi.org/10.1108/IJCHM-07-2016-0402>
- Meyers, L.S., Gamst, G., Guarino, A.J. (2017). *Applied Multivariate Research 3rd Edition*. Thousand Oaks, CA:SAGE Publications.
- Modern Hire. (2020, January 24). *Modern Hire's banner year in 2019*. <https://modernhire.com/newsroom-article/modern-hires-banner-year-in-2019/>
- Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, 62, 760–775. <https://doi.org/10.1111/ajps.12357>
- Myrick, L. (2021). *How to style your room for an online video meeting*. Lesley Myrick Interior Design. <https://lesleymyrick.com/how-to-style-your-room-online-video-meeting/>
- Naim, I., Tanveer, Md. I., Gildea, D., & Hoque, M. E. (2018). Automated analysis and prediction of job interview performance. *IEEE Transactions on Affective Computing*, 9(2), 191–204. <https://doi.org/10.1109/TAFFC.2016.2614299>

- Nestler, S., & Back, M. D. (2013). Applications and extensions of the lens model to understand interpersonal judgments at zero acquaintance. *Current Directions in Psychological Science*, 22(5), 374–379. <https://doi.org/10.1177/0963721413486148>
- Neumann, R., & Strack, F. (2000). Approach and avoidance: The influence of proprioceptive and exteroceptive cues on encoding of affective information. *Journal of Personality and Social Psychology*, 79(1), 39–48. <https://doi.org/10.1037/0022-3514.79.1.39>
- Park, J. H., Schaller, M., & Crandall, C. S. (2007). Pathogen-avoidance mechanisms and the stigmatization of obese people. *Evolution and Human Behavior*, 28(6), 410–414. <https://doi.org/10.1016/j.evolhumbehav.2007.05.008>
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, 42(1), 185–227. <https://doi.org/10.1080/00273170701341316>
- Pryor, J. B., Reeder, G. D., Yeadon, C., & Hesson-McInnis, M. (2004). A dual-process model of reactions to perceived stigma. *Journal of Personality and Social Psychology*, 87(4), 436–452. <https://doi.org/10.1037/0022-3514.87.4.436>
- Raney, J. (2021). *New! Blur background, Zoom phone power pack and expanded availability in 46 countries*. Zoom Blog. <https://blog.zoom.us/new-blur-background-zoom-phone-power-pack-expanded-availability/>
- Roets, A., & Van Hiel, A. (2007). Separating ability from need: Clarifying the dimensional structure of the Need for Closure Scale. *Personality and Social Psychology Bulletin*, 33(2), 266–280. <https://doi.org/10.1177/0146167206294744>

- Roets, A., & Van Hiel, A. (2011). Item selection and validation of a brief, 15-item version of the Need for Closure Scale. *Personality and Individual Differences, 50*(1), 90–94.
<https://doi.org/10.1016/j.paid.2010.09.004>
- Roulin, N., & Bhatnagar, N. (2021). Examining the impact of applicant smoking and vaping habits in job interviews. *Human Relations, 74*(8), 1211–1239.
<https://doi.org/10.1177/0018726720912320>
- Ruggs, E. N., Hebl, M. R., Walker, S. S., & Fa-Kaji, N. (2014). Selection biases that emerge when age meets gender. *Journal of Managerial Psychology, 29*(8), 1028–1043.
<http://dx.doi.org/10.1108/JMP-07-2012-0204>
- Ryan, S., Oaten, M., Stevenson, R. J., & Case, T. I. (2012). Facial disfigurement is treated like an infectious disease. *Evolution and Human Behavior, 33*(6), 639–646.
<https://doi.org/10.1016/j.evolhumbehav.2012.04.001>
- Signs.com. (n.d.). *The influence of video call backgrounds*. <https://www.signs.com/zoom-backgrounds/>
- Stevens, C. K., & Kristof, A. L. (1995). Making the right impression: A field study of applicant impression management during job interviews. *Journal of Applied Psychology, 80*(5), 587–606. <https://doi.org/10.1037/0021-9010.80.5.587>
- Suen, H.-Y., Chen, M. Y.-C., & Lu, S.-H. (2019). Does the use of synchrony and artificial intelligence in video interviews affect interview ratings and applicant attitudes? *Computers in Human Behavior, 98*, 93–101. <https://doi.org/10.1016/j.chb.2019.04.012>
- Summers, J. K., Howe, M., McElroy, J. C., Ronald Buckley, M., Pahng, P., & Cortes-Mejia, S. (2018). A typology of stigma within organizations: Access and treatment effects. *Journal of Organizational Behavior, 39*(7), 853–868. <https://doi.org/10.1002/job.2279>

- 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*(5), 625–638. <https://doi.org/10.1037/apl0000077>
- Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63*(3), 107–140.
<https://doi.org/10.1016/j.cogpsych.2011.06.001>
- Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128*(2), 237–251.
<https://doi.org/10.1016/j.cognition.2012.09.012>
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(3), 520–522. <https://doi-org.library.smu.ca/10.1038/381520a0>
- 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.
<https://doi.org/10.1016/j.ijhm.2018.03.011>
- 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. <https://doi.org/10.1016/j.ijhm.2016.10.012>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185*(4157), 1124–1131.

- UCLA: Statistical Consulting Group. (2021). *Multiple regression power analysis: G*Power data analysis examples*. <https://stats.idre.ucla.edu/other/gpower/multiple-regression-power-analysis/>
- Vasel, K. (2020, March 16). *The rules of videoconferencing at home*. The Mercury News. <https://www.mercurynews.com/2020/03/16/the-rules-of-video-conferencing-at-home/>
- Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, 67(6), 1049–1062. <https://doi.org/10.1037/0022-3514.67.6.1049>
- Wong, O. (2020). *How asynchronous video interview design affects applicant outcomes: Interview performance, impression management, anxiety, and perceived fairness* [Unpublished master's thesis]. Saint Mary's University.

Appendix A: Job Description/Posting

Business Operations Manager – Remote Work Position

Booker's Financial is a digital personal and small business banking company. Our goal each day is to provide simple and convenient banking solutions that promote our clients' financial growth. We deliver tailored solutions both digitally and over the phone to best allow our customers to reach their financial goals, anytime and anywhere. By embracing change, forward-thinking, and the strengths of our diverse and inclusive team, we ensure that our customers are able to remotely access secure and reliable support for all their financial needs.

The Business Operations Manager will assess, recommend, and implement operational objectives and procedures in order to maximize efficiency and support the objectives and long-term growth of the company. This remote work position will involve responsibilities related (but not limited) to policy formation and implementation, budgeting and finances, development of business strategy, and risk assessment. The Business Operations Manager will lead a departmental team working remotely from locations across Canada and the United States, and will liaise with team leads from other related units.

Core Responsibilities:

- Align and optimize operational processes to ensure efficiency and support for strategic corporate objectives
- Develop strategic management initiatives; offer input and guidance on decisions regarding expansion or cost containment
- Regularly liaise with team members from Marketing, Business Development, and Management to devise and/or implement new plans, policies, procedures, and systems
- Identify and manage the risks/impacts associated with the introduction of new solutions, policies, legislation and new business opportunities
- Establish internal and external key performance indicators for the business, and monitor and track the progress of established goals
- Contribute Operations information and recommendations, and assist in the development of business strategy, long range plans, company goals, and growth objectives
- Prepare briefing notes, reports, and presentation material for senior management

Required Knowledge, Skills, Abilities and Other Attributes:

- Proficiency in Microsoft Office and the ability to pick up new technology and video-conferencing (i.e. Microsoft Teams) software easily
- Ability to develop partnerships, collaborate, and communicate with all levels of the organization remotely
- Project-management and organization; ability to coordinate multiple projects with different deadlines, deliverables, and key stakeholders
- Leadership, coaching, and mentoring; ability to support staff and foster employee development remotely

Appendix B: Candidate Response Scripts

Candidate Introduction Video Script:

Hello, my name is Mark Smith, and I am applying for the business operations manager position at Booker's Financial. The competition number is 26959.

High and Low Quality Candidate Response Scripts:

Note. The high quality response scripts are designed to elicit scores of 4 or 5 out of a maximum of 5 points. The low quality responses are designed to elicit scores of 1 or 2 out of a maximum of 5 points.

Question 1: Describe a time when you had to collaborate with others to succeed at a task. What was the task you had to accomplish? What made the collaboration successful? What was your role or contribution?

Teamwork / Collaboration Competency: Working with others to achieve a common objective or complete a shared task.

High Quality Response:

At my last job, I was responsible for finding ways to increase the efficiency of service delivery to our clients. So, one issue that needed to be resolved was the long wait times that our customers experienced while waiting for support over the phone. What I did, was I first asked each department head to develop suggestions as to how the company could improve phone-based customer service practices. From there, I basically hosted a team meeting, in which we evaluated different approaches and weighed the possible pros and cons of certain options. I also made sure that each group member expressed their ideas or concerns, and we ultimately produced a list of three top initiatives. [Pause] I thought that it was very effective to do it that way, as the various subject matter experts were able to identify potential obstacles, and find solutions for them. After that, I drafted a plan of action, and as a group, we confirmed that it was realistic, timely and that it would effectively meet the needs of the organization. [Pause] I also made sure that my colleagues' ideas were included into the final plan of action. Thanks to the team's input, I was able to present a well thought-out plan for the new initiatives, which I should mention, were rolled out exactly as planned. In the end, the project turned out to be even more successful than I expected, and I have my colleagues to thank for that.

Low Quality Response

At my last job, I was responsible for finding ways to increase the efficiency of service delivery to our clients. I was supposed to work with some colleagues as a team, but I knew that I would do a better job without them, so I just went ahead with it independently. So, one issue that needed to be resolved was the long wait times that our customers experienced while waiting for support over the phone. What I did, was I first developed some suggestions as to how the company could

improve phone-based customer service practices and built an action plan. [Pause] Although I presented a well thought-out plan for the new initiatives, unfortunately, my colleagues did not roll it out the way that they were supposed to. In the end, the project turned out to be less successful than I expected, but like I said, it was largely due to my colleagues' failure to implement my plan.

Question 2: Describe a situation where you had to evaluate the risks, benefits, and potential outcomes of a decision. For instance, buying something important, investing in something, starting a new project, etc. How did you handle it? And what was the outcome?

Minimizing Errors/Risks Competency: The considerations that one makes in advance of making a decision or engaging in an action to ensure that the potential for risks, errors, or unfavourable outcomes are minimized.

High Quality Response:

Two years ago, I was, uh, tasked with evaluating the possibility of introducing a new employee academic enrichment program. Basically, the new program would cover some or all of the costs of employees' academic courses. What I did was I first identified the advantages of an academic enrichment program. [Pause] For instance, implementing such a program would potentially show employees that the organization supports their professional development, which could improve employee recruitment, retention and career advancement opportunities. I also thought about the possible disadvantages. For example, an academic enrichment program would likely be very expensive, would need a lot of administrative support, and could reduce employee productivity at work if workers have heavy academic workloads. So, with these pros and cons in mind, I examined a number of departmental reports and did a cost analysis to figure out the value of a new program like that. [Pause] I basically found that if the company covered half of the academic program costs up to a maximum of about \$10,000 per employee, the value added outweighed the costs to the company. From there, I came up with a roll-out plan, which included different ways to monitor the success of the program and prevent any fallout to the company. In the end, the program was successfully implemented, and our annual internal survey actually found that it accounted for a 15% improvement in employee morale.

Low Quality Response:

Two years ago, I was, uh, tasked with evaluating the possibility of introducing a new employee academic enrichment program. Basically, the new program under consideration would essentially cover some or all of the costs of employees' academic courses. I decided that an academic enrichment program should be offered to the employees, regardless of all the problems that my colleagues thought the program would bring. I decided that the initiative would be most popular if the company covered all of employees' academic costs, and assumed that the budget would stretch to cover it. From there, the program was immediately launched across the company. [Pause] Although things went well at first, we ran into several problems that made it necessary to cancel the program. For example, many employees took courses in fields unrelated

to our organization, and the overall costs grossly exceeded the budget. Employees were, uh, not happy about the cancellation of the program, but it was necessary.

Question 3: Describe a time when you took the lead on a group project. What was the project, how did you behave as a leader, and what was the outcome?

Leadership Competency: Coordinating and managing the completion of tasks or objectives among a group of people.

High Quality Response:

A few years ago, I lead a team tasked with developing and reporting a series of recommendations aimed at improving the diversity of our organization. To tackle this project, first I hosted a team meeting. I basically started the meeting by describing the goals of the project, the timeline to complete it, the required deliverables, and the overall impact of the project. From there, I asked the group to, uh, share their diversity ideas. [Pause] Some ideas were good, but many were unrealistic. I made a point of acknowledging all of the ideas and encouraged everyone to participate, but also outlined some of the shortcomings as diplomatically as I could. Some of the quieter members of the group did not share their ideas, so I asked specifically for their opinions and gave them an opportunity to speak. [Pause] At the end of the first meeting, I scheduled a follow-up meeting, and outlined the tasks that needed to be completed prior to that date. After that, I divided the tasks among the group members based on their current workload, strengths and availability. I also made sure that everyone knew that I was the main point of contact should they have any issues moving forward. In the end, all of the members of my group were ready with high quality materials by the next meeting, which allowed me to compile the report well before the deadline set by senior management. It was great to work with such a highly motivated team.

Low Quality Response:

A few years ago, I lead a team tasked with developing and reporting a series of recommendations aimed at improving the diversity of our organization. To tackle this project, first I hosted a team meeting. I basically started the meeting by describing the consequences that team members would face if they did not contribute to the project enough. From there, I asked the group to, uh, share their diversity ideas. Some ideas were good, but many were unrealistic. I made a point of acknowledging the better ideas, while discouraging useless suggestions. [Pause] Some of the quieter members of the group did not share their ideas, so I made a point of telling them that sitting in silence was not helpful. At the end of the meeting, I briefly mentioned the tasks that needed to be completed as soon as possible. After that, I assigned tasks to the group members, and reiterated that the work needed to be completed regardless of their current workload, strengths and availability. I also made sure that everyone knew that they needed to resolve any issues that they encountered on their own, as I was very busy. [Pause] In the end, none of the members of my group were ready with high quality materials by the next meeting. This meant that I was unable to compile the report in time for the deadline set by senior management. I think

that the project would have been more successful if my team had been more highly motivated though.

Question 4: Imagine you've been hired for the position, and you are in your first week on the job. The Business Operations tasks and projects that were previously handled by other staff are now being handed off to you. As a result, you currently already have five ongoing projects, each requiring about 4-5 hours of work to be completed, including two projects with deadlines by the end of the week. Your boss has just given you a new project today, that was described as "important". On top of this, you are still in the process of familiarizing yourself with the company's financial reporting systems, company policies, etc. and find that you have more work than you can comfortably manage. What would you do?

Planning, Prioritization Competency: How one manages and completes multiple different tasks and objectives within a limited time frame, based on an evaluation of each objective's relative importance.

High Quality Response:

First off, I would communicate with my supervisor to determine where the new important task fell with the rest of my prioritized tasks. I would, uh, also find out when the deadline was for that new task, so that I could plan better. [Pause] Next, I would basically create a plan to ensure that the different deliverables were ready in time for each of their deadlines. Like, if it was not possible to complete all of the projects in the amount of time that I had left, I would reach out to my boss with two possible plans: one with the expected timelines if I was able to delegate certain tasks to others, and the other if I did not have that kind of support. And, depending on my boss's answer, I would probably notify all pertinent stakeholders immediately of any expected delays. After that, I would focus on completing the necessary tasks, by prioritizing the ones that needed to be done by the end of the week. For the time being, I would focus only on learning the policies and reporting system information that was needed for those priority tasks. After that, I would build some time into my schedule to make that I learned the policies and other information later on.

Low Quality Response:

First off, I would tell my supervisor that I felt overwhelmed by the unreasonable amount of work that I had to do. Also, I would, uh, ask my supervisor to delegate the new important task to someone else, because it would be asking too much of me at that time. After that, I would start completing the easier tasks regardless of their deadlines, since I would be more familiar with those ones. [Pause] After that, I would just try to get as much work done as I could before the deadlines, and hope that it worked out. Since it would only be my first week, I don't think the organization could expect much more of me.

Question 5: Imagine that it's been a few months since you implemented a new company-wide initiative, aimed at changing the progress reporting system for long-term projects. You are realizing that it is not working the way you had anticipated. A lot of money was spent training employees to use the reporting system, but the compliance rate to your new procedures has been very low. Overall, you have seen no gains in project efficiency, and the upper management team is becoming increasingly concerned. Additionally, you and the management team have heard a few direct complaints from employees on parts of the new system. What would you do? What would you tell the management team?

Problem-Solving Competency: Recognizing when a problem has occurred, identifying the problem, and proposing/implementing an appropriate solution.

High Quality Response:

First off, it would be important to evaluate the problem in more detail. I would probably, uh, contact senior management to propose an assessment strategy in order to figure out what exactly the cause of the problem was. After that, I would reach out to the project leads to learn more about their projects' progress, what challenges they were experiencing, and when they expected to complete those projects. I would ask these questions to essentially determine if other unexpected situational factors may have affected the efficiency of the system. [Pause] If there were no situational factors at work, next I would probably hold employee focus groups, in order to determine if the issue might be more at the worker level. In the end, I would analyze the information that I got from management and the employees, and would integrate it into an overall plan to improve the initiative. From there, I would present my detailed action plan to senior management.

Low Quality Response:

First off, it would be important to allow the situation to run its course without any extra interference from either management or myself. If the senior management team, uh, specifically requested my input to resolve the issue, I would simply reassure them that the problematic elements would improve. After that, I would recommend waiting a few more months, so that the issues could sort themselves out with time. [Pause] If for some reason, the issues persisted, I would at that point probably suggest that perhaps the new initiative should be removed. Ultimately, the decision to keep, improve or discontinue the initiative would rest with senior management though, so it would be more up to them to solve the problem and decide what to do.

Appendix C: Background Manipulations

Professional Background Condition



Personal Background Condition



Blurred Background Condition

