

Singapore Management University

## Institutional Knowledge at Singapore Management University

---

Research Collection Lee Kong Chian School Of  
Business

Lee Kong Chian School of Business

---

4-2024

### The chemistry between us: Illuminating complementarity patterns in interpersonal role play assessment via moment-to-moment analyses

Christoph N. HERDE  
*Singapore Management University*

Filip LIEVENS  
*Singapore Management University, filiplievens@smu.edu.sg*

Follow this and additional works at: [https://ink.library.smu.edu.sg/lkcsb\\_research](https://ink.library.smu.edu.sg/lkcsb_research)



Part of the [Applied Behavior Analysis Commons](#), [Industrial and Organizational Psychology Commons](#), and the [Organizational Behavior and Theory Commons](#)

---

#### Citation

HERDE, Christoph N. and Filip LIEVENS. The chemistry between us: Illuminating complementarity patterns in interpersonal role play assessment via moment-to-moment analyses. (2024). *Journal of Applied Psychology*. 109, (4), 534-550.

Available at: [https://ink.library.smu.edu.sg/lkcsb\\_research/7273](https://ink.library.smu.edu.sg/lkcsb_research/7273)

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [cherylds@smu.edu.sg](mailto:cherylds@smu.edu.sg).

**The Chemistry Between Us: Illuminating Complementarity Patterns in Interpersonal Role  
Play Assessment via Moment-to-Moment Analyses**

Christoph N. Herde

Singapore Management University

Filip Lievens

Singapore Management University

***In press Journal of Applied Psychology***

Both authors contributed equally to this paper. This research was supported by a research grant from Research Foundation - Flanders (FWO), application number: G092512N and by the Ministry of Education, Singapore under its Academic Research Funding Tier 2 (MOE2019-T2-1-191).

We thank Sofie Ameloot, Ellen Bagdasarian, Robin Boudry, Sigrid Lefevre, Chrissie Polfliet, Lisa Van der Schueren, Afra Vanhalst, and Nefert Viaene as well as all participants, role-players, assessors, and coders for their contribution to the data collection. We also thank Frederik Anseel, Ronald Bledow, Gary Burns, Kenneth Tai, and Leigh Mellish for valuable feedback on earlier versions of this paper.

Portions of this paper are based on the doctoral dissertation of Christoph N. Herde and were presented at the 33<sup>rd</sup> Annual Conference of the Society of Industrial and Organizational Psychology (SIOP) and the 81<sup>st</sup> Annual Meeting of the Academy of Management (AoM).

Address of correspondence: Christoph N. Herde. E-Mail: [c.herde@gmx.net](mailto:c.herde@gmx.net)

### Abstract

In assessment and selection, organizations often include interpersonal interactions because they provide insights into candidates' interpersonal skills. These skills are then typically assessed via one-shot, retrospective assessor ratings. Unfortunately, the assessment of interpersonal skills at such a trait-like level fails to capture the richness of how the interaction unfolds at the behavioral exchange level within a role-play assessment. This study uses the lens of interpersonal complementarity theory to advance our understanding of interpersonal dynamics in role-play assessment and their effects on assessor ratings. Ninety-six MBA students participated in four different flash role-plays as part of diagnosing their strengths and weaknesses. Apart from gathering assessor ratings and criterion measures, coders also conducted a fine-grained examination of how the behavior of the two interaction partners (i.e., MBA students and role-players) unfolded at the moment-to-moment level via the Continuous Assessment of Interpersonal Dynamics (CAID) measurement tool. In all role-plays, candidates consistently showed mutual adaptations in line with complementarity principles: Affiliative behavior led to affiliative behavior, whereas dominant behavior resulted in docile, following behavior and vice versa. For affiliation, mutual influence also occurred in that both interaction partners' temporal trends in affiliation were entrained over time. Complementarity patterns were significantly related to ratings of *in situ* (role-playing) assessors but not to ratings of *ex situ* (remote) assessors. The effect of complementarity on validity was mixed. Overall, this study highlights the importance of going beyond overall ratings to capture behavioral contingencies such as complementarity patterns in interpersonal role-play assessment.

**Keywords:** interpersonal complementarity theory; assessment center; role plays; continuous rating assessments; asynchronous video formats

**Running Head:** COMPLEMENTARITY PATTERNS

## **The Chemistry Between Us: Illuminating Complementarity Patterns in Interpersonal Role Play Assessment via Moment-to-Moment Analyses**

Engaging in successful interpersonal interactions and building good relationships are considered to be pivotal for performance in many jobs. To assess interpersonal skills such as communication, influencing others, or consideration of others (Arthur et al., 2003; Huffcutt et al., 2001), organizations have therefore since long invested in selection procedures that feature interactions (e.g., employment interviews, role-plays, or leaderless group discussions; Dayan et al. 2008; Oliver et al., 2016). Candidates also favorably perceive these selection procedures due to their two-way communication (Hausknecht et al., 2004).

However, in selection and assessment, interpersonal skills have typically been assessed at a trait-like level via one-shot assessor/self-ratings of interpersonal behavior at the end of the interaction (e.g., Dayan et al., 2002; Lievens & Klimoski, 2001; Oliver et al., 2016; Waldman et al., 2004). The assessment of interpersonal skills at such a broader, trait-like level fails to capture the richness of how interactions unfold within, for instance, role-plays or interviews. Accordingly, we lack a solid understanding of what actually happens at the detailed level of interpersonal behaviors and interpersonal dynamics (i.e., changes in the interpersonal behavior of one party due to the behavior of the other party during an interaction). The need to better understand the moment-to-moment changes in the behavior of interaction partners becomes especially evident when one aims to shed light on dynamic phenomena. Interpersonal complementarity constitutes such a critical dynamic phenomenon. It transpires in everyday social contexts (e.g., friendship, romantic relationships) and refers to “the dynamic interplay of behavior between two people during interaction - that is, people’s mutual adjustments to each other, and therefore the changes in behavior, that occur during the course of interaction” (Sadler et al., 2009, p.1006). According to interpersonal complementarity theory (Carson, 1969; Kiesler, 1983), these

adjustments follow functional rules: (un)friendly behavior leads to (un)friendly behavior, whereas dominant behavior results in submissive behavior and vice versa<sup>1</sup>.

Although interpersonal complementarity serves as a governing force of daily interactions, we have no evidence whether it also occurs in interpersonal assessment: Do candidates adapt to their interaction partners in accordance with the complementarity rules? Or do they breach these rules, such as reacting dominantly when exposed to dominant behavior? And how do such patterns unfold over the course of the interaction? Moreover, it is concerning that we are in the dark as to whether potential complementarity patterns affect the ratings that assessors give to candidates. So, do candidates receive lower/higher evaluations when their interactions breached or followed the rules of complementarity? Another pressing question is whether these complementarity patterns represent diagnostic (criterion-relevant) information.

Therefore, the objective of this paper is twofold. We seek to better understand the occurrence of complementarity patterns in interpersonal assessment situations (i.e., can complementarity patterns be discerned in interpersonal assessments?). In addition, we examine whether these complementarity patterns affect the ratings that assessors provide and these ratings' validity for predicting future performance. Due to the dynamic nature of complementarity we conduct a fine-grained investigation of how the interactions continuously unfold at the moment-to-moment level of behavioral exchange within a given assessment situation (Gabriel et al., 2017; Jebb & Tay, 2017; Krasikova & LeBreton, 2012)<sup>2</sup>. In our

---

<sup>1</sup> In this study, we use the terms dominance and affiliation, although agency and communion are also used in the literature. As noted by Moskowitz et al. (2007), dominant or agentic behaviors include “expressing an opinion, taking the lead in planning or organizing, waiting for the other person to act first, and not expressing disagreement” (p.1053), whereas affiliative or communal behaviors refer to “agreeable and quarrelsome behaviors, such as listening attentively to the other, compromising about a decision, making a sarcastic comment, and intentionally providing incorrect information (p.1053).

<sup>2</sup> Generally, interpersonal behavior can be described at three levels: (a) an overall tendency across time and interaction partners, (b) an overall tendency in a specific situation, and (c) momentary behavior over time in a situation. Although all levels can provide valuable insights, the continuous behavioral exchange level is best suited to uncover interpersonal dynamics (e.g., Hopwood et al., 2020; Markey et al., 2010; Sadler et al., 2009; Tracey, 2004).

investigation, we focus on one popular and pivotal interpersonal assessment method, namely role-plays (in the form of speeded or flash role-plays, Herde & Lievens, 2020; 2023).

This study makes several important contributions. First, we contribute to a better understanding of the interpersonal microcosmos in role-play assessment by focusing on the interplay between candidates and role-players. As a key advantage, our momentary (continuous) measurement permits addressing behavioral contingencies between two interdependent parties (role-players and candidates; Oostrom et al., 2019), compared to past studies that have relied one-shot, retrospective assessor/self-ratings of interpersonal behavior. Generally, continuous measurement generates a more precise understanding of how the role-play unfolds at the behavioral exchange level and whether specific configurations can be distinguished. Second, we draw on interpersonal complementarity theory (Carson, 1969; Kiesler, 1983) in our hypothesis development to introduce the principles of complementarity as potential drivers underlying such configurations. Hereby we distinguish between mutual “influence” (overall temporal trend) and mutual “adaptations” (continuous adjustments; Cappella, 1996). Third, we go beyond examining the occurrence of complementarity patterns in role-play assessment but also hypothesize about their consequences for candidates (do they affect the ratings they are given?) and organizations (do they affect the criterion-related validity of these ratings?).

## **Study Background**

### **Interpersonal Complementarity Theory**

Interpersonal complementarity theory (Carson, 1969; Kiesler, 1983; see also Abele & Wojciszke, 2007; Moskowitz & Zuroff, 2005; Pincus & Ansell, 2013) proposes that interpersonal behavior can be described in a Cartesian plane with two orthogonal bipolar axes (i.e., the interpersonal circumplex). The x-axis describes the dimension of affiliation (also known as communion) that covers a continuum of behaviors expressing warmth or friendliness (positive

side) vs. behaviors expressing coldness or hostility (negative side). The y-axis covers behaviors expressing dominance (also known as agency) or leading others (positive side) vs. behaviors expressing submissiveness or following others (negative side). Substantial empirical evidence confirmed that the underlying structure of interpersonal behavior can be described via these affiliation and dominance dimensions. Similarly, there is support that this organization of interpersonal behavior follows a circular structure (Markey et al., 2003; Moskowitz, 1994; Tracey et al., 2001; Figure 1).

Interpersonal complementarity theory also aims to explain and predict interpersonal dynamics in everyday situations (Carson, 1969; Kiesler, 1983). The general notion is that each behavior of one actor rewards the prior behavior and constrains the subsequent behavior of the other actor. In this process, these interpersonal dynamics are proposed to follow prescribed patterns: Actors adjust their behavior to be compatible with their counterparts along two complementarity principles (rules). The first principle relates to affiliation. For example, in an interaction between Ann and Lisa, interpersonal complementarity theory assumes that when Ann displays friendly behavior, Lisa will correspond with friendly behavior. The second principle of complementarity is related to dominance: It is assumed that when Ann displays dominant (leading) behavior, Lisa will respond with docile (following) behavior.

According to interpersonal complementarity theory, these patterns develop at a subtle and instantaneous level (see Markey et al., 2010; Tiedens & Fragale, 2003; Tiedens et al., 2007). In addition, the theory posits that adherence to these complementarity principles is functional because it constitutes a pivotal foundation for successful cooperation in everyday situations (Carson, 1969; Kiesler, 1983; Pincus & Ansell, 2013; Wiggins, 1980). Showing complementarity means that both interaction partners accept their status in the hierarchy, thereby contributing to a process of self-validation and feelings of security. Yet, when the principles are violated, the

interaction partners are expected to experience interpersonal anxiety and a less effective collaboration.

Finally, a temporal perspective can be added to these interpersonal patterns. As argued by Cappella (1996), complementarity can be observed as an overall temporal trend (mutual “influence”) and at the behavioral exchange level (mutual “adaptations”). In *mutual influence*, the continuous complementarity patterns accumulate in an overall *temporal trend*. Figure 2a shows how over the course of the interaction, two parties adjust their dominance behavior in line with the complementarity principle (their dominance trends run opposite to each other), whereas Figure 2b displays how over the course of the interaction two parties adjust their affiliative behavior in line with the complementarity principle (their affiliation trends are entrained).

Conversely, Cappella (1996) referred to *mutual adaptations* as a “dynamic process by which partners respond to changes in one another’s behavior during interaction” (p.354). So, the focus is on the occurrences in which one party changes its interpersonal behavior in response to the other party. That is why mutual adaptations are denoted as specific *momentary adjustments* in one party’s behavior that are linked to changes in the other party’s behavior, holding overall temporal trends constant (Sadler et al., 2009, Sadler et al., 2011). Figure 3a and 3b depict mutual adaptations in terms of dominance and affiliation, respectively. As can be seen, incidents of dominant behavior by one interaction partner are followed by obeying behaviors by the other one, whereas the opposite occurs for affiliation.

## Hypotheses

### Complementarity Patterns in Interpersonal Role Play Assessment

Although there exists a vast amount of research on interpersonal complementarity theory, the two complementarity principles have so far been predominantly documented in everyday (nonwork) situations. Hence, less is known whether they also play out in an assessment



environment. As compared to most nonwork settings, interpersonal role-play assessment simulates workplace interactions, is more structured, and invokes maximum performance due to the short time period and evaluative context (Ployhart et al., 2001): Candidates tend to put their best foot forward by trying to read the situation (Bangerter et al., 2012; Jansen et al., 2013; Kleinmann, 1993). Therefore, we do not know whether in interpersonal assessment the complementarity principles would still account for a substantial portion of candidates' interactional behavior.

Two prior studies (Moskowitz et al., 2007; Oliver et al., 2016) shed initial light on this issue. Moskowitz et al.'s (2007) diary study examined how context (nonwork/leisure vs. work) affects the occurrence of complementarity. They found stronger evidence for dominance complementarity in work than nonwork settings. Conversely, affiliation complementarity was stronger in nonwork settings. According to Moskowitz et al., dominance complementarity fits better with the structured (often problem solving-related) work situations, whereas affiliation complementarity thrives in unstructured, nonwork situations. Despite this study's insights on complementarity in work settings, it did not deal with interpersonal role-play assessment. Oliver et al. (2016) conducted the only study about complementarity patterns in role-plays (as part of a medicine program). Yet, they used retrospective self-reports as a crude approach for assessing dominance and affiliation complementarity. Similar to Moskowitz et al., dominance complementarity was stronger in this context. Role-players' overall affiliation and candidates' overall relationship-building were even negatively related.

Considering these results, we expect the interpersonal dynamics between candidates and role-players to follow the dominance complementarity principle because role-plays confront candidates with problems and require them to propose solutions. In fact, when role-players actively and more dominantly lead the discussion via verbal (e.g., ask targeted questions, propose

solutions, make demands) and nonverbal behavior, Interpersonal Complementarity posits candidates will display more docile (non)verbal behavior (e.g., listen to gather all available information) to solve the problem. Conversely, when role-players are more passive (e.g., ask for help, do not disclose information), interpersonal complementarity theory states candidates to lead the discussion via their (non)verbal behavior (e.g., with targeted questions to gain relevant information or propose solutions). Therefore, in role-plays, dominance complementarity should typically be instrumental for solving problems. As compared to dominance complementarity, we expect affiliation complementarity (i.e., either friendly-friendly or hostile-hostile exchanges at the behavioral level) to be a less prominent force in role-play assessment: Although reacting to the interaction partner by showing corresponding (non)verbal behavior (i.e., friendly-friendly or hostile-hostile) might either improve or hamper relationship building, it is less directly related to better problem solving (see also Oliver et al. 2016).

On the basis of these empirical and conceptual arguments, we posit that in interpersonal assessments dominance complementarity will govern more strongly interactions between candidates and role-players than affiliation complementarity. We further hypothesize these complementarity patterns to occur at the level of momentary behavioral exchanges (mutual adaptations) and then to accumulate as an overall temporal trend (mutual influence). Thus:

*Hypothesis 1 (H1a): Complementarity at the overall temporal trend level (indicative of mutual influence) in role-plays is stronger for dominance than for affiliation.*

*Hypothesis 1 (H1b): Complementarity patterns at the momentary level (indicative of mutual adaptations) in role-plays are stronger for dominance than for affiliation.*

### **Complementarity Patterns and Their Effect on Assessors' Evaluations of Candidates**

Above we put forward a first set of hypotheses related to complementarity patterns in interpersonal assessment. The next hypotheses assume that there are individual differences in

these patterns. That is, we expect that complementarity patterns will occur more in the behavioral exchanges of some candidates than in those of other candidates. A key unexplored question is then whether such complementarity patterns in turn have an effect on assessor evaluations of candidates. We posit the type of assessor to play a pivotal role here. Along these lines, Rauthmann and Sherman (2020) make a distinction between *in situ*, *juxta situm*, and *ex situ* assessors. *In situ* assessors are physically present and involved in the interaction. So, they personally experience and “sense” the interaction (partner). For example, an *in situ* rater is a role-player that also serves as assessor. *Juxta situm* assessors are present in the assessment environment but they are observers (bystanders) and are thus not involved in the interaction. These are the typical assessors in brick-and-mortar assessment centers (ACs). *Ex situ* assessors are neither part of the assessment environment nor personally involved in it because they evaluate off-line/remotely (via video). In this study, we focus on comparing *in situ* and *ex situ* raters because there were no *juxta situm* assessors. Using *ex situ* raters (instead of *juxta situm* raters) is relevant because remote assessors are increasingly used due to the growth of online assessment (Mercer, 2018, p.29; Pinsight, 2019).

According to the assessment and selection literature, assessors are thoroughly trained to focus on behavior, use rating aids, and avoid rating effects (e.g., Lievens, 1998; Lievens, Schollaert et al., 2015; Roch et al., 2012). This leads to the expectation that complementarity patterns should not affect assessors’ evaluations. We posit that this will indeed be the case but only for *ex situ* assessors. Recall that interpersonal complementarity theory posits that complementarity patterns occur at a subtle, instantaneous level (Markey et al., 2010; Tiedens & Fragale, 2003; Tiedens et al., 2007). Although complementarity patterns might evolve out of verbal exchanges, they thus also result from subtle changes in nonverbal and paralinguistic behavior such as changes in eye contact, smiling, physical proximity, and posture. Research

shows that such complementarity signals make the interaction inherently satisfying so that both interaction partners like each other more (Sadler et al., 2011). Although remote assessors observe the same verbal behavior as *in situ* assessors, they are not involved in the interaction and thus we posit that they are less “exposed” to the nonverbal and paralinguistic exchanges between the two parties and the resulting rewarding effects (e.g., being confronted oneself with (non)verbal rude/friendly behavior is different from watching someone else facing such behavior in a movie). Indeed, computer-mediated interactions are typically seen as inferior in terms of “social presence” (eye contact, smiling, physical proximity, and posture; Short et al., 1976; Van Iddekinge et al., 2006). Conversely, complementarity patterns might affect *in situ* assessors’ ratings of the candidates and *in situ* assessors might implicitly factor in these ‘rewarding’ aspects in their ratings.

So, we expect that complementarity patterns (both mutual influence and adaptations) are significantly related only to candidate evaluations of *in situ* assessors. This expectation echoes one of the reasons why it is suggested that role-players should not simultaneously serve as assessors, even though so far this recommendation has remained untested. Thus:

*Hypothesis 2 (H2a): Complementarity patterns at the overall temporal trend level (indicative of mutual influence) are positively related to in situ assessors’ evaluations of candidates, whereas this is not the case for ex situ assessors.*

*Hypothesis 2 (H2b): Complementarity patterns at the momentary level (indicative of mutual adaptations) are positively related to in situ assessors’ evaluations of candidates, whereas this is not the case for ex situ assessors.*

### **Complementarity Patterns and Their Effect on the Validity of Assessors’ Evaluations**

Although the prior section addressed whether complementarity patterns affect assessors’ evaluations of candidates, a last critical unexplored issue is: Do these complementarity patterns

affect the validity of these evaluations for predicting future real-life performance? Our expectation is that complementarity patterns in interpersonal role-play assessment represent criterion-relevant variance. This expectation is based on empirical research in social and clinical psychology that supports beneficial outcomes of interpersonal complementarity: It was positively related to satisfaction with the interaction (Locke & Sadler, 2007; Tracey, 2004), liking (Markey et al., 2010; Tiedens & Fragale, 2003), and relationship quality (Ansell et al., 2008; Markey & Markey, 2007). Apart from these subjective outcomes in nonwork contexts, empirical research also documented the advantageous effects of complementarity on performance in lab tasks (Estroff & Nowicki, 1992; Markey et al., 2010) and at work (Grant et al., 2011; Hu & Judge, 2017; Wilson et al., 2016). For example, Grant et al. (2011) found that the effect of leaders' dominance (extraversion) on team performance was moderated by employees' proactivity level: Teams with employees voicing their suggestions produced lower performance with more extraverted leaders, whereas the opposite occurred for teams with employees who did not voice alternative courses of action.

Conceptually, complementarity originates from behavior that is couched in a context, namely in an ongoing sequence of interpersonal behavior of how one party reacts to the other one. Although so far these behavioral contingencies have been neglected in assessment contexts, Oostrom et al. (2019) recommended explicitly assessing such sophisticated interpersonal dynamics because they reflect how interactions also unfold in actual job situations. So, a precise assessment of how role-plays unfold at the behavioral exchange level between two parties and the resulting complementarity configurations provide assessors with more "context" when they observe behavior and they might factor this into their ratings, resulting in higher validity.

In sum, these empirical and conceptual arguments suggest that in work settings complementarity patterns (both mutual influence and adaptations) lead to better performance and

that in assessment situations these patterns thus represent systematic criterion-relevant variance.

Thus, we posit the following hypothesis:

*Hypothesis 3 (H3): Complementarity patterns (both in terms of mutual influence and adaptations) in role-plays represent criterion-relevant variance.*

## Method

### Sample and Procedure

The sample comprised of the entire MBA cohort of a European business school, namely 96 candidates (51% females, mean age = 23.63,  $SD = 1.85$ ) from 19 different countries. All had at least one year of work experience as a junior manager and had chosen either an MBA major in Marketing or Financial Management. They participated in an assessment program to identify their strengths and weaknesses. Hence, candidates' mean test motivation was high: 3.96 ( $SD = 0.50$ )<sup>3</sup>, as measured via the scale of Arvey et al. (1990; four items with the anchors: 1 = *strongly disagree*; 5 = *strongly agree*; alpha = .67). Their high involvement was also supported by anecdotal evidence. For example, candidates wore business attire and were nervous. We excluded one candidate from the analyses because she did not take part in the role-plays.

To set up this assessment program, we collaborated with a consultancy firm. One week prior to the role-plays, candidates completed proctored computer-based tests (e.g., personality and cognitive ability). Candidates completed brief role-plays in which they interacted with different role-players. The sequence in which they participated in the role-plays followed a carousel-like procedure, with the first role-play for each candidate being randomly determined to limit order effects. Candidates' performance during each role-play was rated by the designated role-playing assessor (*in situ* assessor) and independent remote assessors (*ex situ* assessors via

---

<sup>3</sup> This value is based on  $n = 49$  because unfortunately test motivation data for about half of the sample were not registered.

video recordings in which pausing or rewinding were prohibited because this was also not possible for role-playing assessors)<sup>4</sup>. In line with best practices for online assessment, we asked the remote assessors to create a standardized rating environment (a silent place without distraction and they had headphones on). Most of them (80%) did their ratings in a proctored computer lab of the university. One week after the role-plays, candidates received feedback reports. Seven months later, MBA program instructors rated them to provide criterion data. Finally, a distinct sample of coders watched the videos and rated both candidates' and role-players' moment-to-moment interpersonal behavior.

### **Role-plays**

Given the interaction partner is the main situational feature in role-plays it is important to test the occurrence of complementarity patterns not only in one role-play but across several role-plays that vary in terms of role-player behavior (script). Due to this need to include more than one role-play, we decided to focus on speeded ("flash") role-plays. Recently, given cost/efficiency concerns and the pandemic, such multiple, speeded assessments have made inroads into practice under various formats (Herde & Lievens, 2020; Herde & Lievens, 2023). Examples are online scenario-based simulations (wherein candidates respond via a webcam to prerecorded video scenes in which an actor speaks into the camera, Mercer, 2017, 2018) and online role-plays (which allow back-and-forth interaction between a remote role-playing assessor and a candidate, Byham, 2016; Pinsight, 2019)<sup>5</sup>. In our study, we decided to go for role-plays with a duration of three minutes because (1) previous interpersonal research demonstrates that

---

<sup>4</sup> To reduce order effects, we distributed all records for the two role-plays per assessor across four blocks that each contained records of only one role-play, counterbalanced the records across assessors, and randomly presented candidates to them. For 22 % of all conversations, assessors used audio records to evaluate performance due to camera recording issues.

<sup>5</sup> Although speeded assessment formats are a recent employee selection trend, there is a long tradition in healthcare (certification and admission) to use Objective Structured Clinical Examinations (Brannick et al., 2011; Harden et al., 1975) and Multiple Mini Interviews (Eva et al., 2004; Knorr & Hissbach, 2014), which share many features with multiple, speeded assessment.

three minutes are long enough for complementarity patterns to potentially occur (e.g., Hopwood et al., 2020) and (2) prior social and personality psychology research (“thin slices paradigm”) shows that three minutes is enough for providing accurate judgments (e.g., Ambady & Rosenthal, 1992; Carney et al., 2007; Murphy et al., 2015).

As there does not exist a situational taxonomy in ACs (see Hoffman et al., 2015 for a broad coding scheme for AC exercise situations) and the interaction partner is the key situational factor in role-plays, it was a logical decision to draw from interpersonal complementarity theory to choose the role-plays’ main underlying situational features. As noted, this theory states that the affiliation and dominance of the situation (i.e., interaction partner) are two key features to be considered. Crossing these two features leads to a quadrant with four basic situations. Hence, our role-plays represented the four quadrants of the interpersonal circumplex (see Table 1). Inclusion of these four role-plays with varying situational demands (each of the four quadrants) ensures that we did not choose a role-play situation that might have favored complementarity patterns to occur. Accordingly, it allows drawing more robust, generalizable conclusions that are not contingent upon one role-play situation. For these reasons, prior studies (e.g., Fournier et al., 2008; 2009) have opted for the same design to examine complementarity in everyday situations.

We took these four role-plays from a set of 18 flash role-plays (that were part of a larger study, see Herde & Lievens, 2023) that we constructed together with the consultancy firm. As inspiration for these speeded role-plays, the consultancy firm relied on longer role-plays that they had designed in the past. From our part, we ensured that the role-plays were realistic via input from experienced consultants and assessment experts. Note that all role-plays were framed in an over-arching theme (the organization of a charity event). The candidate played the role of project manager and faced different people (portrayed by role-players) from inside and outside the organization. Each role-player put forward an issue related to the charity event.



These four specific role-plays were taken because each of them matched one quadrant of the interpersonal circumplex. For example, as shown in Table 1, one role-play (role-play 2) represents the low affiliation and high dominance quadrant because the role-player was unfriendly and criticized the candidate (low affiliation). The role-player also urged the candidate to take a decision (high dominance). According to interpersonal complementarity theory, this means that the candidate in turn will display low affiliative behavior and low dominance behavior. This assumption is not evident because in an assessment context one might expect that the candidate also engages in socially desirable behavior and thus shows the opposite behavior, namely behavior high in affiliation (see Oliver et al., 2016). Table 1 describes the other three role-plays in their respective quadrants. A pilot study (see results in the Appendix) confirmed that each role-play reflected its respective quadrant.

### **Role-players, Assessors, Rating Process, and Ratings**

A total of 13 role-players (12 females) acted in the role-plays. They were either consultants or Master students from a large European university. None of them were acquainted with the MBA students. All role-players attended a training session (Byham, 1977; Lievens, Schollaert et al., 2015) wherein they were familiarized with the materials (background script, prompts, etc.), practiced their role, and received feedback about their role-playing. Role-players were randomly assigned to one role-play. As noted, each of them was also a live (in situ) assessor. The paid *ex situ* (remote) assessors (4 females, 4 males; mean age = 23.13,  $SD = 6.49$ ) were students from the same university as the role-playing assessors.

All (*in situ* and *ex situ*) assessors received the same behavior-driven (Byham, 1977) and frame-of-reference training (Roch et al., 2012). It included lectures and exercises on observation, registration, classification, and evaluation. In addition, assessors were familiarized with the

overall scoring procedure. Next, they independently practiced evaluating three videotaped candidate performances in the role-plays they were specialized in.

All assessors used the same rating form to provide two or three overall ratings of performance after each role-play. An example is: 1 = “*should clearly be improved: starters’ level*” to 9 = “*obviously strong: role model performance*”. In case of flash role-plays, use of such overall ratings has been recommended (e.g., Brannick, 2008; Lievens, 2008). To ensure that ratings were based on observable and relevant behaviors, we developed checklists with behaviors indicative of (in)effective performance (e.g., *assesses the consequences of the different alternatives*). Across role-plays and role-players/assessors, the average internal consistency reliability of these ratings was .70. We computed a composite of *in situ* assessors’ ratings per role-play and a composite of *ex situ* assessors’ ratings per role-play. Note further that the composite of *ex situ* assessors’ ratings was an aggregate of all available *ex situ* assessors’ ratings.

To estimate the interrater reliabilities of the ratings, we calculated a  $G(q,k)$  coefficient, which is an ICC for ill-structured measurement designs (Putka et al., 2008). The mean  $G[q,k]$  for performance ratings averaged across all assessors per role-play was .65, which is typical for interrater reliability in such contexts (Herde & Lievens, 2020).

### **Moment-to-moment Assessment of Interpersonal Behavior**

**CAID.** We used the Continuous Assessment of Interpersonal Dynamics (CAID; Sadler et al., 2009) to measure candidates’ and role-players’ momentary interpersonal behavior in each role-play. CAID works with a joystick (Sadler et al., 2009) whose x-axis represents the affiliation continuum, with a scale from -1000 (extreme unfriendliness) to +1000 (extreme friendliness). The y-axis reflects the dominance continuum, with the scale from -1000 (extreme submissiveness) to +1000 (extreme dominance). In the CAID, coders watch a recorded dyadic interaction and focus on one target at the time. They monitor the target’s interpersonal behavior,

thereby moving the joystick to the position within the interpersonal circumplex that represents the current level of affiliation and dominance. The software displays the interpersonal circumplex and the joystick's current position on the screen and continuously (every 0.5 second) writes data points for both affiliation and dominance (Sadler et al.). Prior research documents the adequate reliability and construct-related validity of CAID codings (see also below).

The real-time snapshots of interactions at the behavioral exchange level via CAID capture individuals constantly providing each other with subtle cues about their dominance and affiliation. Accordingly, the CAID is a key improvement over prior approaches that coded interpersonal behavior within a speaking turn (e.g., Strong et al., 1988) because important interpersonal dynamics might also occur between speaking turns (Hopwood et al., 2020).

**Coders, training, and coding.** A total of 17 Bachelor (59 %) or Master students in Psychology from a European university (16 females, mean age = 21.67,  $SD = 1.35$ ) were paid to serve as coders. They received an 8-hour training session to learn the CAID. Similar to past studies (Markey et al., 2010; Sadler et al., 2009; Sadler & Woody, 2016), the training consisted of two parts: The first part aimed to develop a common frame-of-reference of affiliation and dominance. After a lecture on the interpersonal circumplex and the background of the role-plays, coders were familiarized with 9-point BARS (Oliver, 2012) indicative of different degrees of affiliation and dominance. They then independently practiced rating overall affiliation and dominance via these BARS in six videotaped role-plays, discussed their ratings, and received feedback. The second part focused on applying the CAID (Sadler & Woody, 2016). After a lecture on the CAID, the trainer mentioned adjectives indicative of the quadrants and coders had to move the joystick to the corresponding quadrant. It was emphasized that joystick ratings had to be based on changes in a target's relevant, observable behaviors (Sadler et al., 2009). Afterwards, coders watched how the trainer coded a target. Finally, coders practiced the CAID in at least

seven practice tapes. The trainer monitored the practice codings and presented means, standard deviations, cross-correlations as well as plots of the codings for feedback purposes.

All videotapes of the role-plays were distributed to four or five coders. Similar to other CAID studies, a coder coded both candidate and role-player in two distinct runs. Given a fixed camera angle had been used, coders were instructed to use all information about targets' interpersonal behavior at any moment (see Sadler et al., 2009, Sadler et al., 2015). When targets were not fully visible, targets still presented auditory cues (e.g., sighs, "uh-huhs") or showed body movements (e.g., leaning into the camera). When targets were not speaking, not presenting auditory cues, or not changing their nonverbal behavior, coders were instructed to keep the joystick in the same position. The same instructions were given to coders of audio files for interactions not successfully recorded on video. Although the use of CAID on audio files loses visual and nonverbal information, prior interpersonal research (especially in psychotherapy) shows the CAID approach has been successfully used to uncover complementarity patterns on the basis of audio files (see e.g., Tracey, 2012). In addition, our pilot studies revealed that CAID codings of audio files and video files of the same vignettes led to similar results (e.g., similar mean CAID scores and similar levels of adaptation in dominance and affiliation). Details of these pilot studies are reported in the online supplement. In total, coders coded for 286 hours. To limit order effects, tape order was randomized.

**Mutual influence and mutual adaptation measures.** The CAID produces two bivariate time series (one for affiliation levels and one for dominance levels over time) for both the candidate and role-player. Each time series has about 360 data points because codings were made every 0.5 second for 3 minutes. In prior studies, this time lag of 0.5 seconds was determined as ideal (e.g., Sadler et al., 2009). Similar to prior studies, we omitted codings related to the first ten

data points (i.e., first five seconds) because coders need to settle into the interaction (e.g., Sadler et al., 2009; Thomas et al., 2014).

To obtain mutual influence and mutual adaptation scores per candidate, we followed the approach outlined by Sadler et al. (2009) and decomposed the bivariate timeseries (i.e., two timeseries: one for the candidate and one for the role-player) by running in each timeseries regressions per dimension (affiliation or dominance) and role-play. In all regressions, we predicted the time series of affiliation or dominance of each party by time in seconds as predictor. Accordingly, we decomposed the affiliation (or dominance) time series of each party into (1) an intercept, (2) a slope and (3) a time series of residuals. According to Sadler et al., the intercept indicates the level of affiliation (or dominance) of a party at the start. The slope (b coefficient) indicates any overall temporal trend in affiliation (or dominance) of a party across the duration of the interaction. Finally, the time series of residuals indicate dynamic changes in affiliation (or dominance) of a party, with the temporal trend across the interaction being controlled for. As both parties were unacquainted at the start, the intercepts in their respective regressions did not provide useful information in this study<sup>6</sup>. Yet, in the regressions of the timeseries, the entrainment of the slopes (indicative of mutual influence) and entrainment of the residuals (i.e., indicative of mutual adaptations) enabled testing our hypotheses.

Specifically, the entrainment (cross-correlation) between candidates' and role-players' slopes (b coefficients) of the dominance time-series served as a measure of mutual dominance influence, whereas the entrainment (cross-correlation) between candidates' and role-players' slopes (b coefficients) of the affiliation time-series served as a measure of mutual affiliation influence. To capture the entrainment of slopes of both parties' time series per dyad, we used

---

<sup>6</sup> That said, the role player intercepts are generally in line with the imposed role-player disposition (interpersonal circumplex quadrant). For example, in both role-play 1 and 2, the initial level of affiliation of the role player was low (see also Appendix).

formulas like Ansell et al. (2008) did: Mutual influence affiliation =

$\sqrt{(b \text{ candidate} - b \text{ role player})^2}$ ); mutual influence dominance =

$\sqrt{(b \text{ candidate} + b \text{ role player})^2}$ ). These mutual influence scores indicate deviations from perfect complementarity. To align the interpretation of mutual influence for affiliation with the common interpretation of this effect in complementarity theory, we reversed the sign so that higher values for affiliation influence reflect higher affiliation complementarity. Consistent with complementarity theory, lower values for dominance influence reflected higher complementarity.

The entrainment (cross-correlation) between candidates' and role-players' time series residuals of the dominance time-series served as a measure of mutual dominance adaptations, whereas the entrainment (cross-correlation) between candidates' and role-players' time series residuals of the affiliation time-series served as a measure of mutual affiliation adaptations. The interpretation of affiliation and dominance adaptations is the same as above. Additional information on the distribution of mutual influence and adaptation scores can be found in the online supplement.

### **Other Measures**

**Social Behavior Inventory.** To check the construct-related validity of CAID codings, role-players assessed candidates' overall affiliation and dominance via selected items from the Social Behavior Inventory (SBI; Moskowitz, 1994) after each role-play. Prior empirical studies revealed that the SBI generates reliable and valid measures of affiliation and dominance (for an overview, see Moskowitz & Sadikaj, 2012). Role-players indicated whether behavior related to the interpersonal circumplex (i.e., agreeableness, quarrelsomeness, dominance, and submissiveness) was shown by the candidates, not shown, or not applicable. All scales had acceptable internal consistency reliabilities (average across all role-plays from .69 to .85).

We computed candidates' overall affiliation on the SBI per role-play by subtracting mean quarrelsomeness scores from mean agreeableness scores (Moskowitz, 1994). Overall dominance per role-play was computed by subtracting mean submissiveness scores from mean dominance scores. We then averaged scores across role-plays.

### **Criterion Measures**

Seven months after the role-plays, each candidate was rated by the MBA program instructors. During the MBA, action-based learning modules and workshops required candidates to work in self-managed groups, draft reports, and give presentations. Candidates also worked together for three months on a real-life consultancy project. Instructors often addressed candidates' progress and problems, gave advice on how to solve problems, and provided feedback. So, instructors had ample opportunity to observe and assess the candidates.

To obtain criterion data we relied on the relative percentile method (Carver et al., 2021; Goffin et al., 1996; Goffin et al., 2009). In this method, a rater is asked to compare the performance of an individual to a reference group. The rater is told that the reference group consists of the average "employee" (i.e., a percentile of 50). Although this method is less well known than absolute rating methods, there is consistent evidence that the relative percentile method leads to psychometric advantages over traditional absolute ratings. In particular, the relative percentile method has been proven to be effective in reducing rating inflation in supervisor/self-ratings (see Goffin & Olson, 2011). Note that the relative percentile method differs from forced distribution ratings because the rater is not "forced" to place a percentage of rates in specific categories.

Using this relative percentile method, raters assigned percentile scores to ratees per criterion dimension. Instructors rated four criterion dimensions: Two of those were related to adapting one's interpersonal behavior (team member adaptivity, interpersonal adaptability),

whereas the other two were related to task performance (task-oriented leadership, in-role behavior). In the rating form, each dimension was described via anchors of established scales (Griffin et al., 2007; Mumford et al., 2007; Williams & Anderson, 1991; Yukl, 1999).

To investigate the structure underlying criterion ratings, we ran confirmatory factor analyses in Mplus 7.4 (ML estimator). The relative percentile method ratings served as indicators. We compared a one-factor model (Model A) with a model with two correlated factors of task performance and interpersonal adaptability (Model B). Only Model B showed a good fit with exception of a poor RMSEA (Model A/B:  $\chi^2(df) = 35.17(2)/1.62(1)$ ,  $p < .001$ / $p = .203$ , CFI = .769/.996, RMSEA (90% CI) = .418 (.304-.544)/.081 (.000-.299), SRMR = .087/.014). Information criteria also indicated Model B to better fit than Model A (Model A/B: AIC: 3413.94/3382.39). So, we averaged instructor ratings into two criteria. Alpha was .81 for both task performance and interpersonal adaptability. The two factors correlated .50 ( $p < .001$ ).

### **Transparency and Openness**

We describe our sampling plan, all data exclusions (if any) and all measures identified for this study, and we adhered to the *Journal of Applied Psychology* methodological checklist. The data are part of a larger study (Herde & Lievens, 2023; Hickman, Herde, Lievens, & Tay, 2023; Lievens, Corstjens, & Herde, 2023). Analysis data and code are available from the corresponding author upon reasonable request. Given this study was a collaboration with an international consultancy firm, some research materials (role-plays, scoring sheets, training materials) are not available due to their proprietary nature, whereas others (test motivation scale and criterion measures) are included in the Online Supplement. Data were analyzed using R, version 3.6.3 (R Core Team, 2015) and the *psych* package version 2.0.12 (Revelle, 2020) as well as Mplus 7.4. The study design was not preregistered because the data were collected in the context of an applied assessment project. The hypotheses and analysis were not preregistered because at that



time this was not common practice. The institutional review board of Ghent University approved this study (IRB protocol number 2020/13).

## **Analyses and Results**

### **Preliminary Analyses: Reliability and Construct-related Validity of CAID codings**

We calculated interrater reliabilities of the CAID codings via cross-correlations between coders' time series codings across coders. Similar to Markey et al. (2010), we used the Spearman–Brown prediction formula to compute the four-coder reliability. The values were .53 for affiliation and .69 for dominance. These results are similar to past studies (e.g., Dermody et al., 2017; Fox et al., 2021; Markey et al., 2010; Sadler et al., 2009; Thomas et al., 2014) and thus we also aggregated the codings for each target, per time point, across the four coders.

To examine the convergent validity of CAID codings, we correlated candidates' average affiliation and dominance CAID codings with role-players' ratings on these two dimensions via the SBI. Across all role-plays, convergent validities with the SBI ratings were  $r = .41$  ( $p < .001$ ) and  $r = .44$  ( $p < .001$ ), for CAID affiliation and dominance codings respectively. We examined discriminant validity by computing the correlation between the average affiliation CAID coding and the dominance SBI rating and the correlation between the average dominance CAID coding and the affiliation SBI rating. Across all role-plays, discriminant validities with the SBI ratings were  $r = .11$  ( $p = .266$ ) and  $r = .02$  ( $p = .843$ ) for CAID affiliation and dominance codings respectively. Thus, in line with prior research (Sadler et al., 2009), there was evidence of both convergent and discriminant validity of the CAID codings.

### **Is Complementarity Present in Interpersonal Dynamics in Speeded Role-plays?**

We started by inspecting the bivariate time series of each candidate-role player dyad. As examples, Figure 4 depicts the bivariate time series related to one dyad's momentary levels across one entire role-play in this study. In Figure 4a, the two parties often follow the dominance

complementarity principle because they almost simultaneously change their dominance level in opposing direction to the other. In Figure 4b, there are also many instances of both parties almost simultaneously changing their affiliation level in the same direction, although the size of the entrainment (correlation) of time series in Figure 4b is lower than in Figure 4a.

To go beyond visually inspecting time series, a straightforward approach to quantify the degree of mutual adaptation in moment-to-moment responses consists of computing a correlation between parties' scores over time (as done in Figure 4a and 4b). Unfortunately, such cross-correlations might also be prone to confounds. Most critically, they might represent mainly a shared overall linear trend instead of similarity of patterns across time in bivariate time series (Cappella, 1996; Sadler et al., 2009). To avoid this issue, Sadler et al. outlined a specific analytical strategy to decompose the time series data, which we adopted (see Method).

H1a proposed that the temporal trends in affiliation and dominance levels over the course of the role-play would follow complementarity principles (and thus be indicative of mutual influence) but that the trend in dominance levels would be larger than the one in affiliation levels. Given the slope of each party's regression indicates such a linear trend over the course of the role-play, a significant relation between these trends of the interaction partners then indicates mutual influence (Cappella, 1996; Sadler et al., 2009). To investigate mutual influence in dominance complementarity we thus calculated the correlation between candidates' and role-players' slopes of the dominance time-series (see Sadler et al., 2009 and see *Method*). The sign of this correlation should be negative because dominance complementarity assumes an opposite trend in dominance among the parties. Across all role-plays, all correlations were indeed negative and the average correlation equaled  $-.17$  ( $SD = 0.09$ ).

To investigate mutual influence in affiliation complementarity we did the same and thus computed the correlation between candidates' and role-players' slopes of the affiliation time-

series (Sadler et al., 2009). Here the sign should be positive because this principle posits a tendency among the parties toward the same behavior. Results showed that candidates' and role-players' overall trend in affiliation levels followed the complementarity principle. Across all role-plays, the average correlation was positive and equaled .32 ( $SD = 0.12$ ).

In sum, these results do not support H1a because there was somewhat more evidence for mutual influence in affiliation complementarity (.32) than for mutual influence in dominance complementarity (-.17). Inspection of the results for the individual role-plays confirmed this. For dominance complementarity, the negative correlation between both parties' slopes was significant in only two role-plays: 3 ( $r = -.28, p = .007$ ) and 4 ( $r = -.21, p = .039$ ), but not in role-plays 1 ( $r = -.07, p = .471$ ) and 2 ( $r = -.13, p = .209$ ). For affiliation complementarity, significant positive correlations between both parties' slopes emerged in three role-plays: 1 ( $r = .31, p = .002$ ), 3 ( $r = .45, p < .001$ ), and 4 ( $r = .33, p = .001$ ). The correlation was not significant in role-play 2 ( $r = .16, p = .126$ ).

H1b posited that mutual adaptations in dominance complementarity would be larger than mutual adaptations in affiliation complementarity. To investigate whether the parties adapted their momentary dominance expressions to each other we correlated candidates' and role-players' residuals of their dominance time series (see Sadler et al., 2009 and *Method*). Table 2 presents the results. As Pearson correlations have a skewed sampling distribution and Table 2 displays averages of correlations, the presented correlations were transformed to the Fisher's z metric (Corey et al., 1998; Hopwood et al., 2020; Silver & Dunlap, 1987). We found consistent evidence of mutual adaptations because the average size of the correlations between the residuals was -.58 and there were significant negative correlations in all four role-plays (Table 2). This indicates, for example, that role-players' submissive expressions evoke dominant ones from candidates and vice versa.

To investigate mutual adaptations in affiliation complementarity, we computed correlations between candidates' and role-players' residuals of their affiliation time series. The average size of the correlations equaled .25. This evidence was also consistent because there were significant positive correlations in all four role-plays (Table 2).

In sum, these results support H1b. Although complementarity patterns were present for mutual adaptations in both dominance complementarity and affiliation complementarity in all four role-plays, the evidence for the former was indeed stronger (-.58) than for the latter (.25).

### **Do Complementarity Patterns Affect Assessors' Evaluations of Candidates?**

The above analyses reveal that at the behavioral exchange level the interpersonal dynamics between candidates and role-players follow complementarity principles for dominance and affiliation in 4 out of 4 role-plays, indicating mutual adaptation. For affiliation, there was also evidence of complementarity at the overall trend level, indicating mutual influence. As we discovered that the interpersonal dynamics in role-plays can be described along the axioms of interpersonal complementarity theory, our second set of hypotheses tested whether these complementarity patterns affect the ratings that *in situ* assessors gave to candidates (as compared to *ex situ* assessors).

To test H2a, we calculated correlations between mutual influence (entrainment of slopes of both parties' time series) and assessor ratings (either *in situ* or *ex situ*). We tested H2a with ratings aggregated across all role-plays. In line with H2a, only for *in-situ* assessors, significant correlations were found. The correlation between affiliation complementarity and *in situ* assessors' ratings was positive (.21,  $p = .040$ , see Table 3). So, when both parties' affiliation levels were positively entrained over the course of the role-plays, *in situ* assessors' ratings were higher. As expected, the sign of the correlation between dominance complementarity and *in situ* assessors' ratings was negative (-.22,  $p = .033$ ). Thus, opposite shifts in dominant behavior across

the role-play were rewarded with higher ratings. Conversely, for *ex situ* assessors, correlations were .14 ( $p = .166$ ) for affiliation and  $-.04$  ( $p = .712$ ) for dominance. This supports H2a.

To test H2b, we calculated correlations between mutual adaptations (entrainment of residuals of both parties' time series) and assessors' ratings (either *in situ* or *ex situ*). Again, we found significant correlations only for *in situ* assessor ratings. Mutual adaptations related to affiliation complementarity correlated .21 ( $p = .036$ ) with *in situ* assessors' ratings, whereas momentary changes related to dominance complementarity correlated  $-.26$  ( $p = .010$ ). For *ex-situ* assessor ratings, the correlations were .13 ( $p = .216$ ) for affiliation and  $-.14$  ( $p = .160$ ) for dominance. This lends support to H2b.

### **Do Complementarity Patterns Affect Criterion-related Validity?**

Our last hypothesis tackled the relationship between complementarity patterns and candidates' performance seven months later. Table 3 shows that both *in situ* and *ex situ* assessor ratings had validities above .30, attesting to the good criterion-related validity of the flash role-plays. H3 posited that complementarity patterns contain criterion-relevant information. We tested H3 via a variety of analyses. First, we correlated mutual influence and mutual adaptations (for both affiliation and dominance) with candidates' criterion ratings seven months later. Seven out of eight correlations were in the expected direction and two were significant: Mutual influence in affiliation complementarity correlated significantly with interpersonal adaptability ( $.23, p = .025$ ), whereas mutual adaptations in dominance complementarity correlated significantly with task performance ( $-.22, p = .031$ ).

Second, we calculated the partial correlation between *in situ* assessor ratings and criterion performance, thereby controlling for complementarity patterns (i.e., mutual influence and mutual adaptations for both affiliation and dominance). If such complementarity patterns represent error in assessor ratings, controlling for them will make the validity coefficient larger. If these patterns

capture criterion-relevant variance, controlling for them will make the validity coefficient smaller. Results showed that when controlling for complementarity the validity of *in situ* assessor ratings remained constant (remaining at .34,  $p = .001$ , for predicting task performance) or slightly decreased (from .31,  $p = .003$  to .27,  $p = .009$ , for predicting interpersonal adaptability).

In sum, results related to H3 were mixed. There was some relationship between interpersonal complementarity and criterion performance, although the validity correlations were small. In addition, there was no evidence that interpersonal complementarity harmed the validity of assessor ratings. But, at the same time, there was also little or no evidence that such complementarity explained a substantial portion of the validity of these ratings.

### **Additional Analyses**

We ran additional analyses to check whether the complementarity patterns found might be explained via shared variance with mean CAID scores and still add incremental variance beyond these mean CAID scores. The correlations between complementarity patterns and mean CAID scores varied between  $-.36$  and  $.24$ . We also reran all our analyses, thereby controlling for mean CAID scores. Conclusions remained the same (apart from mutual adaptations in dominance also affecting *ex situ* assessors' ratings). So, the effects of complementarity did not disappear when we controlled for mean CAID levels. Detailed results can be found in the online supplement.

## **Discussion**

### **Main Conclusions and Implications for Theory**

This study deepened our knowledge about the dynamics in interpersonal role-play assessment. Although candidates and role-players might engage in a wide variety of interactional patterns, a first key discovery was that these configurations could be described to some extent along the axioms of interpersonal complementarity theory, namely via the principles of complementarity in affiliation and complementarity in dominance. We found cross correlations

of  $-.58$  for mutual adaptations in dominance complementarity and  $.25$  for mutual adaptations in affiliation complementarity. The size of these correlations and the difference between dominance and affiliation complementarity echo those in previous studies in everyday situations, even though our evidence for affiliation is somewhat less pronounced (Hopwood et al., 2020; Markey et al., 2010; Sadler et al., 2009). These results are noteworthy because in assessment settings, candidates might show a variety of interactional behavior (i.e., they are not “instructed” to follow complementarity principles). Our role-plays were brief encounters between unacquainted people. Similar to common AC practices but contrary to everyday situations (Moskowitz et al., 2007), trained role-players were also used. Although role-player training might reduce complementarity (which as a dyadic phenomenon is also dependent on role-players), this did not occur.

That said, the fact that the correlations are not higher and thus that complementarity explains the interactional configurations only to some extent signals that candidates also show other interactional patterns. Candidates might deviate from the complementarity rules to influence and nudge the interaction in a desired direction. For instance, they might mirror competitive role-player responses (contrary to dominance complementarity) or remain friendly when challenged (and thus not display correspondence in low affiliation). Our somewhat lower evidence for mutual adaptations in affiliation shows that in this assessment context such conscious efforts from candidates to shift the conversation in a specific direction and thus to breach complementarity patterns occur more frequently for affiliation than for dominance. Interpersonal complementarity theory does not exclude such exchanges. However, the key point is that this theory posits that the two complementarity principles (and especially dominance) explain significant portions of the variance in candidate-role-player exchanges. And this is what we found consistently across *all four* situations of the interpersonal circumplex.

Second, we went beyond describing the occurrence of complementarity patterns in general. Candidates vary in the extent that they show complementarity patterns with the role-player. Therefore, it was important to link these patterns also to outcome criteria that matter for candidates and companies. Both mutual influence and mutual adaptations affected the ratings of *in situ* assessors (role-playing assessors). Conversely, the effects on the criterion-related validity of the ratings of these *in situ* assessors were mixed. Complementarity did neither significantly increase nor significantly decrease validity. That said, it was striking that the *temporal trend of affiliation complementarity* predicted an interpersonal outcome. This underscores the importance of selecting candidates who can align their affiliation levels with their interaction partner across the whole role-play. Conversely, *dominance complementarity adaptations* predicted a task outcome: When candidates can swiftly adapt their dominance expressions to the interaction partner they also get the task done on the job. To the best of our knowledge, we are the first to discover such differential relationships for mutual influence (affiliation) and mutual adaptations (dominance). So, they add to current theorizing on interpersonal complementarity theory.

Third, this study highlights the importance of going beyond *overall* complementarity when testing the axioms of interpersonal complementarity theory. Past research's inconclusive evidence for complementarity patterns (Oliver et al., 2016) was likely due to the less appropriate reliance on single-point, retrospective assessor ratings or on self-ratings of interpersonal behavior. Conversely, our continuous, behavioral exchange level approach is better aligned with the logic of interpersonal complementarity theory. This might explain why these complementarity patterns consistently occurred and explained significant portions of the variance in *in situ* assessors' ratings.

Fourth, this study introduced the concepts of mutual influence and mutual adaptations in this domain. This distinction disentangles *trends* over the course of an event from within-event



*adaptations*. Whereas the former indicate how overall slopes of the time series of both persons are entrained with each other and are conceptualized as mutual “influence”, the latter reflect moment-to-moment changes in one party’s behavior in response to the other party and are considered mutual “adaptations” (Cappella, 1996; Sadler et al., 2009). We found consistent evidence for mutual adaptations, whereas results indicative of mutual influence were less evident. Given this was an assessment setting, it would have been plausible for mutual influence to occur less in affiliation (see Oliver et al., 2016) because this means, for example, that candidates show a temporal trend towards low affiliation across the course of the role-play when being faced with a low affiliation role-play(er). However, this was not the case because there was evidence for mutual influence for affiliation in three of the four role-plays. For dominance, mutual influence occurred in half of the role-plays. As a possible explanation, the number of mutual adaptations might have not been large enough to aggregate into a significant mutual influence trend because candidates also deviated from the principles of complementarity (see above). Relatedly, three minutes might have been too short for mutual influence to unfold (e.g., Nowicki & Manheim, 1991; Markey & Kurtz, 2006).

Fifth, our results underscore the need to account for behavioral contingencies between two parties in terms of interpersonal patterns in role-plays. Moreover, we showed the value of conceptualizing the myriad of interpersonal configurations along the propositions of interpersonal complementarity theory. Our study thus leverages the notion that interpersonal assessments are an interactive dialogue in which both parties’ behavior is couched in a context, namely in an ongoing sequence of other interpersonal behavior (Ostrom et al., 2019). Both parties adapt their behavior to the other party. So, more broadly, our results call for adopting a more inclusive, contextualized perspective in the study of interpersonal behavior in recruitment and assessment encounters.

### **Avenues for Future Research**

Momentary (real-time) measures surpass past studies' reliance on single point, retrospective self-reports (Oliver et al., 2016) to study interpersonal dynamics. Hence, these continuous measurements open a new era of interpersonal research. As Gabriel et al. (2017) posited: "if ESM produces data like a photo album and diary methods produce short movie summaries, then CRA [continuous rating assessment] provides an entire (albeit brief) movie" (p.34). So, future research should extend our fine-grained approach to other recruitment and assessment encounters by scrutinizing mutual influence and mutual adaptations between job seekers and recruiters/interviewers. In these encounters, affiliation complementarity patterns at the behavioral exchange level might account for rapport building. So far, rapport building has been conceptualized as a holistic notion and thus remains poorly understood (Swider et al., 2016; Swider et al., 2022). Interpersonal dynamics and leader emergence in group simulations also deserve attention (Gerpott et al., 2019; Lehmann-Willenbrock & Allen, 2014). For example, dominance complementarity patterns (in mutual adaptations and mutual influence) might signal that one of the candidates gradually emerge as a leader. As a last example, our approach could be used to scrutinize conflict management strategies (Tjosvold et al., 2014). These examples show how the theoretical lens of interpersonal complementarity and our aligned methodology might lead to a more detailed understanding of these important phenomena.

As a second and related avenue, future studies should examine complementarity patterns in tandem with impression management tactics (Klehe et al., 2014; McFarland et al., 2005). Connecting these research streams permits to answer whether complementarity mediates the effect of impression management (e.g., ingratiation) on performance ratings. Along these lines, we should also scrutinize other possible mediators such as liking of the interaction partner.

Again, adopting a real-time, behavioral exchange level approach might deepen our understanding of the interplay of these processes.

Third, future studies could further unpack the residual variance in the bivariate time series to discover other forms of interpersonal complementarity such as cycles in both parties' interpersonal behavior or interdependent oscillations (Sadler et al., 2011). This recommendation is especially relevant for longer role-plays or leaderless group discussions. To this end, more advanced statistical techniques (e.g., cross-spectral analyses) could also be explored (Vowels et al., 2021; Zhou et al., 2021).

### **Limitations**

First, we note some generalizability caveats. Our results are based on MBA students and on only four role-plays. These role-plays were chosen to reflect varying situational demands (interpersonal circumplex quadrants) so that conclusions did not depend on a specific situation. Yet, future studies should manipulate other situational demands that were analyzed in interpersonal research such as role-player status (subordinate vs. colleague vs. supervisor; Moskowitz et al., 2007; Yao & Moskowitz, 2015). The assessment took place with a developmental focus. So, we welcome replications with actual candidates in other AC simulations. Our criterion ratings also came from MBA instructors. Although they had opportunity to observe the MBA students in various task and interpersonal settings over seven months, future studies should test our hypotheses with supervisors at work.

Second, the short (3-minute) role-plays deserve attention. Our choice of shorter role-plays reflected a recent trend in practice (e.g., Herde & Lievens, 2020; Byham, 2016; Pinsight, 2019; Mercer, 2017, 2018). Prior research and our construct-related validity results document three minutes to be enough for complementarity patterns to be validly captured by the CAID approach (Hopwood et al., 2020). Inclusion of speed role-plays also ensured a strong test because our

evidence shows complementarity patterns occur even in 3-minute structured interactions.

Although one might argue complementarity patterns to flatten out with longer durations, this is unlikely because it runs counter interpersonal research (e.g., Markey et al., 2010; Sadler et al., 2011, 2009). Yet, replication of our results in longer AC exercises is needed.

Third, this study focused only on *in situ* assessors (assessors that are involved in the role-plays and assessment) and *ex situ* assessors (remote assessors that are not involved in role-plays) who provided overall ratings per role-play. Our study did not include *juxta situm* assessors (bystander assessors that are physically present but are not involved in the role-plays), which are the norm in brick-and-mortar ACs. In addition, remote assessors were prohibited to pause or rewind the videotapes to make their task as similar as possible to the one of the role-playing assessors. Therefore, we call for future studies to extend our results to *juxta situm* (bystander) assessors and other assessor rating conditions.

### **Implications for Practice**

First, our results suggest that organizations should be cognizant about complementarity patterns and their rewarding influence on both interaction partners. To assess candidates' general tendency to adapt their behavior to interaction partners (Griffin et al., 2007; Oliver & Lievens, 2014; Ployhart & Bliese, 2006; Pulakos et al., 2000), organizations can even go a step further and confront candidates with a variety of brief interpersonal situations (with the interpersonal circumplex as guide to sample them) or train role-players to explicitly use prompts that might evoke complementarity patterns. Accordingly, organizations can determine whether candidates can establish complementarity patterns. This might be a better approach than assessing adaptability via self-or other-ratings.

Second, our results also speak to the practice of removing the back-and-forth interactional component of role plays by presenting candidates with prerecorded role play vignettes (e.g.,

Cucina et al., 2015; DeSoete et al., 2013; Lievens, DeCorte et al. 2015, Lievens et al., 2019; Oostrom et al., 2010), whereby candidates then react via their webcam to each of these vignettes. Similarly, asynchronous video interviews (e.g., Hickman, et al., 2021) no longer involve back-and-forth exchanges and reciprocity. Apart from this practice eliminating the two-way communication inherent in traditional interviews, our results show that organizations miss out on the possibility to observe whether candidates can establish complementarity with the interaction partner over the course of the interaction.

Third, our results speak to differences between *in situ* (role-playing) vs. *ex situ* (remote) assessors. *Ex situ* assessors are more neutral than *in situ* assessors because ratings from the latter reward the complementarity being established. *Ex situ* assessors also do not face the cognitively taxing dual task of role-playing and evaluating. So, the question is whether organizations want to factor in complementarity in assessor ratings. If so, then *in situ* assessors can be used because complementarity patterns did neither significantly increase nor decrease validity.

Finally, let us clarify that we do not argue for the CAID approach to be used in assessment practice as input for decision making in selection because it is very time intensive. We do believe that the underlying philosophy (i.e., emphasis on moment-to-moment exchanges, dynamics, and behavioral contingencies) provides a window of opportunity for revamping feedback and development interventions. Although feedback typically tends to focus on individual behaviors (Taylor et al., 2005), this study shows the value of including moment-to-moment exchanges and behavioral contingencies. For example, together with a coach, candidates might qualitatively review how their performances unfold on video (i.e., how do they react nonverbally and verbally to role player statements).

## **Conclusion**

This study's fine-grained and real-time investigation obtained unprecedented evidence about what is exactly happening in interpersonal role-plays at the behavioral exchange level. Results were supportive of interpersonal complementarity theory because in all role-plays the momentary adjustments between candidates and role-players could be consistently described along the predicted complementarity patterns for both affiliation and dominance. In addition, complementarity patterns affected the ratings of *in situ* assessors but not those of *ex situ* assessors.

### References

- Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93*, 751–763.  
<https://doi.org/10.1037/0022-3514.93.5.751>
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256–274.  
<http://dx.doi.org/10.1037/0033-2909.111.2.256>
- Ansell, E. B., Kurtz, J. E., & Markey, P. M. (2008). Gender differences in interpersonal complementarity within roommate dyads. *Personality and Social Psychology Bulletin, 34*, 502–512.
- Arthur Jr, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125–153.
- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology, 43*, 695–716.
- Bangerter, A., Roulin, N., & König, C. J. (2012). Personnel selection as a signaling game. *Journal of Applied Psychology, 97*, 719–738.
- Brannick, M. T. (2008). Back to basics of test construction and scoring. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 131–133.
- Brannick, M. T., Erol-Korkmaz, H. T., & Prewett, M. (2011). A systematic review of the reliability of objective structured clinical examination scores. *Medical Education, 45*, 1181–1189.
- Byham, W. C. (1977). Assessor selection and training. In J. L. Moses & W. C. Byham (Eds.), *Applying the assessment center method* (pp. 89–125). New York: Pergamon Press.
- Byham, W. (2016). *Assessment centers for large populations*. Presented at the International Congress on Assessment Center Methods, Bali, Indonesia.
- Cappella, J. N. (1996). Dynamic coordination of vocal and kinesic behavior in dyadic interaction: Methods, problems, and interpersonal outcomes. In J. H. Watt & C. A. Van Lear (Eds.), *Dynamic patterns in communication processes* (pp. 353–386). Thousand Oaks, CA: Sage.
- Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality, 41*, 1054–1072.  
<https://doi.org/10.1016/j.jrp.2007.01.004>
- Carson, R. C. (1969). *Interaction concepts of personality*. Chicago, IL: Aldine.
- Carver, S., Goffin, R., & Factor, R. (2021). Relative and absolute self-ratings of work performance: Does social comparison lead to an enhanced thought process in self-evaluations?. *International Journal of Selection and Assessment, 29*, 336–351.
- Corey, D., Dunlap, W.P., & Burke, M.J. (1998). Averaging correlations: Expected values and bias in combined Pearson's r and Fisher's z transformations. *Journal of General Psychology, 125*, 245–261.
- Cucina, J. M., Su, C., Busciglio, H. H., Harris Thomas, P., & Thompson Peyton, S. (2015). Video-based testing: A high fidelity job simulation that demonstrates reliability, validity, and utility. *International Journal of Selection and Assessment, 23*, 197–209.
- Dayan, K., Kasten, R., & Fox, S. (2002). Entry-level police candidate assessment center: An efficient tool or a hammer to kill a fly?. *Personnel Psychology, 55*, 827–849.
- Dayan, K., Fox, S., & Kasten, R. (2008). The preliminary employment interview as a predictor of assessment center outcomes. *International Journal of Selection and Assessment, 16*, 102–111.

- Dermody, S. S., Thomas, K. M., Hopwood, C. J., Durbin, C. E., & Wright, A. G. (2017). Modeling the complexity of dynamic, momentary interpersonal behavior: Applying the time-varying effect model to test predictions from interpersonal theory. *Journal of Research in Personality, 68*, 54-62.
- De Soete, B., Lievens, F., Oostrom, J., & Westerveld, L. (2013). Alternative predictors for dealing with the diversity–validity dilemma in personnel selection: The constructed response multimedia test. *International Journal of Selection and Assessment, 21*, 239–250.
- Estroff, S., & Nowicki, S. (1992). Interpersonal complementarity, gender of interactants, and performance on puzzle and word tasks. *Personality and Social Psychology Bulletin, 18*, 351–356.
- Eva, K. W., Rosenfeld, J., Reiter, H. I., & Norman, G. R. (2004). An admissions OSCE: The multiple mini-interview. *Medical Education, 38*, 314–326.
- Fournier, M. A., Moskowitz, D. S., & Zuroff, D. C. (2008). Integrating dispositions, signatures, and the interpersonal domain. *Journal of Personality and Social Psychology, 94*, 531–545.
- Fournier, M. A., Moskowitz, D. S., & Zuroff, D. C. (2009). The interpersonal signature. *Journal of Research in Personality, 43*, 155-162.
- Fox, S. E., Thomas, K. M., Durbin, C. E., & Hopwood, C. J. (2021). A variance decomposition of the Continuous Assessment of Interpersonal Dynamics. *Psychological Assessment, 33*, 427–442 .
- Gabriel, A. S., Diefendorff, J. M., Bennett, A. A., & Sloan, M. D. (2017). It’s about time: The promise of continuous rating assessments for the organizational sciences. *Organizational Research Methods, 20*, 32–60.
- Garcia, M. F., Posthuma, R. A., & Colella, A. (2008). Fit perceptions in the employment interview: The role of similarity, liking, and expectations. *Journal of Occupational and Organizational Psychology, 81*, 173–189.
- Gerpott, F. H., Lehmann-Willenbrock, N., Voelpel, S. C., & van Vugt, M. 2019. It’s not just what is said, but when it’s said: A temporal account of verbal behaviors and emergent leadership in self-managed teams. *Academy of Management Journal, 62*, 717–738.
- Goffin, R. D., Gellatly, I. R., Paunonen, S. V., Jackson, D. N., & Meyer, J. P. (1996). Criterion validation of two approaches to performance appraisal: The behavioral observation scale and the relative percentile method. *Journal of Business and Psychology, 11*, 23–33.
- Goffin, R. D., Jelley, R. B., Powell, D. M., & Johnston, N. G. (2009). Taking advantage of social comparisons in performance appraisal: The relative percentile method. *Human Resource Management, 48*, 251–268.
- Goffin, R. D., & Olson, J. M. (2011). Is it all relative? Comparative judgments and the possible improvement of self-ratings and ratings of others. *Perspectives on Psychological Science, 6*, 48-60.
- Grant, A. M., Gino, F., & Hofmann, D. A. (2011). Reversing the extraverted leadership advantage: The role of employee proactivity. *Academy of Management Journal, 54*, 528–550.
- Griffin, M. A., Neal, A., & Parker, S. K. (2007). A new model of work role performance: Positive behavior in uncertain and interdependent contexts. *Academy of Management Journal, 50*, 327–347.



- Harden, R. McG., Stevenson, M., Downie, W. W., & Wilson, G. M. (1975). Assessment of clinical competence using objective structured examination. *British Medical Journal*, *1*, 447–451
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *57*, 639–683.
- Herde, C. N., & Lievens, F. (2020). Multiple Speed Assessments: Theory, practice, and research evidence. *European Journal of Psychological Assessment*, *36*, 237–249.
- Herde, C.N., & Lievens, F. (2023). Multiple, speeded assessments under scrutiny: Underlying theory, design considerations, reliability, and validity. *Journal of Applied Psychology*, *108*, 351-373.
- Hickman, L., Bosch, N., Ng, V., Saef, R., Tay, L., & Woo, S. E. (2021). Automated video interview personality assessments: Reliability, validity, and generalizability investigations. *Journal of Applied Psychology*, early online access.
- Hickman, L., Herde, C.N., Lievens, F., & Tay, L. (2023). Automatic scoring of speeded interpersonal assessment center exercises via machine learning: Initial psychometric evidence and practical guidelines. *International Journal of Selection & Assessment*, *31*, 225-339.
- Hoffman, B. J., Kennedy, C. L., LoPilato, A. C., Monahan, E. L., & Lance, C. E. (2015). A review of the content, criterion-related, and construct-related validity of assessment center exercises. *Journal of Applied Psychology*, *100*, 1143-1168.
- Hopwood, C. J., Harrison, A. L., Amole, M., Girard, J. M., Wright, A. G. C., Thomas, K. M. ... Kashy, D. A. (2020). Properties of the Continuous Assessment of Interpersonal Dynamics across sex, level of familiarity, and interpersonal conflict. *Assessment*, *27*, 40-56.
- Hu, J., & Judge, T. A. (2017). Leader–team complementarity: Exploring the interactive effects of leader personality traits and team power distance values on team processes and performance. *Journal of Applied Psychology*, *102*, 935–955.
- 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*, 897-913.
- Jansen, A., Melchers, K. G., Lievens, F., Kleinmann, M., Brändli, M., Fraefel, L., & König, C. J. (2013). Situation assessment as an ignored factor in the behavioral consistency paradigm underlying the validity of personnel selection procedures. *Journal of Applied Psychology*, *98*: 326-341.
- Jebb, A. T., & Tay, L. (2017). Introduction to time series analysis for organizational research: Methods for longitudinal analyses. *Organizational Research Methods*, *20*, 61–94.
- Kiesler, D. J. (1983). The 1982 interpersonal circle: A taxonomy for complementarity in human transactions. *Psychological Review*, *90*, 185–214.
- Klehe, U.-C., Kleinmann, M., Nieß, C., & Grazi, J. (2014). Impression management behavior in assessment centers: Artificial behavior or much ado about nothing? *Human Performance*, *27*, 1–24.
- Kleinmann, M. (1993). Are rating dimensions in assessment centers transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology*, *78*, 988-993.
- Knorr, M., & Hissbach, J. (2014). Multiple mini-interviews: Same concept, different approaches. *Medical Education*, *48*, 1157–1175.
- Krasikova, D. V., & LeBreton, J. M. (2012). Just the two of us: Misalignment of theory and methods in examining dyadic phenomena. *Journal of Applied Psychology*, *97*, 739–757.

- Lehmann-Willenbrock, N., & Allen, J. A. (2014). How fun are your meetings? Investigating the relationship between humor patterns in team interactions and team performance. *Journal of Applied Psychology, 99*, 1278–1287.
- Lievens, F. (1998). Factors which improve the construct validity of assessment centers: A review. *International Journal of Selection and Assessment, 6*, 141–152.
- Lievens, F. (2008). What does exercise-based assessment really mean? *Industrial and Organizational Psychology, 1*, 112–115.
- Lievens, F., Corstjens, J., & Herde, C.N. (2023). Multiple, speeded assessments: Initial evidence on subgroup differences and applicant perceptions. *International Journal of Selection & Assessment, 31*, 240-251.
- Lievens, F., De Corte, W., & Westerveld, L. (2015). Understanding the building blocks of selection procedures: Effects of response fidelity on performance and validity. *Journal of Management, 41*, 1604–1627.
- Lievens, F., & Klimoski, R. J. (2001). Understanding the assessment center process: Where are we now? *International Review of Industrial and Organizational Psychology*, (vol. 16, pp. 245–286). Chicester: John Wiley & Sons, Ltd.
- Lievens, F., Sackett, P. R., Dahlke, J. A., Oostrom, J. K., & De Soete, B. (2019). Constructed response formats and their effects on minority–majority differences and validity. *Journal of Applied Psychology, 104*, 715-726.
- Lievens, F., Schollaert, E., & Keen, G. (2015). The interplay of elicitation and evaluation of trait-expressive behavior: Evidence in assessment center exercises. *Journal of Applied Psychology, 100*, 1169–1188.
- Locke, K. D., & Sadler, P. (2007). Self-efficacy, values, and complementarity in dyadic interactions: Integrating interpersonal and social-cognitive theory. *Personality and Social Psychology Bulletin, 33*, 94–109.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*, 99–128.
- Markey, P. M., Funder, D. C., & Ozer, D. J. (2003). Complementarity of interpersonal behaviors in dyadic interactions. *Personality and Social Psychology Bulletin, 29*, 1082–1090.
- Markey, P. M., & Kurtz, J. E. (2006). Increasing acquaintanceship and complementarity of behavioral styles and personality traits among college roommates. *Personality and Social Psychology Bulletin, 32*, 907-916.
- Markey, P., Lowmaster, S., & Eichler, W. (2010). A real-time assessment of interpersonal complementarity. *Personal Relationships, 17*, 13–25.
- Markey, P. M., & Markey, C. N. (2007). Romantic ideals, romantic obtainment, and relationship experiences: The complementarity of interpersonal traits among romantic partners. *Journal of Social and Personal Relationships, 24*, 517-533.
- McFarland, L., Yun, G. J., Harold, C. M., Viera, L., & Moore, L. G. (2005). An examination of impression management use and effectiveness across assessment center exercises: The role of competency demands. *Personnel Psychology, 58*, 949–980.
- Mercer (2017). *Talent trends 2017 global study: Empowerment in a disrupted world*. Mercer.
- Mercer (2018, April 19th). *Global talent trends 2018 study: Unlocking growth in the human age*. Mercer: North America Webcast.
- Moskowitz, D. S. (1994). Cross-situational generality and the interpersonal circumplex. *Journal of Personality and Social Psychology, 66*, 921–933.

- Moskowitz, D. S., Ho, M. -h. R., & Turcotte-Tremblay, A.-M. (2007). Contextual influences on interpersonal complementarity. *Personality and Social Psychology Bulletin*, *33*, 1051–1063.
- Moskowitz, D. S., & Sadikaj, G. (2012). Event-contingent recording. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 160–175). New York: Guilford Press.
- Moskowitz, D. S., & Zuroff, D. C. (2005). Assessing interpersonal perceptions using the interpersonal grid. *Psychological Assessment*, *17*, 218–230.
- Mumford, T. V., Campion, M. A., & Morgeson, F. P. (2007). The leadership skills strataplex: Leadership skill requirements across organizational levels. *The Leadership Quarterly*, *18*, 154–166.
- Murphy, N. A., Hall, J. A., Schmid Mast, M., Ruben, M. A., Frauendorfer, D., Blanch-Hartigan, D., Roter, D. L., & Nguyen, L. (2015). Reliability and validity of nonverbal thin slices in social interactions. *Personality and Social Psychology Bulletin*, *41*, 199–213. <https://doi.org/10.1177/0146167214559902>
- Nowicki, S., & Manheim, S. (1991). Interpersonal complementarity and time of interaction in female relationships. *Journal of Research in Personality*, *25*, 95–99.
- Oliver, T. (2012). *Applying a framework of interpersonal adaptability for assessment*. University of Guelph.
- Oliver, T., Hausdorf, P., Lievens, F., & Conlon, P. (2016). Interpersonal dynamics in assessment center exercises: Effects of role player portrayed disposition. *Journal of Management*, *42*, 1992–2017.
- Oliver, T., & Lievens, F. (2014). Conceptualizing and assessing interpersonal adaptability. *Individual adaptability to changes at work: New directions in research* (pp. 52–72). New York: Taylor & Francis.
- Oostrom, J. K., Born, M. P., Serlie, A. W., & van der Molen, H. T. (2010). Webcam testing: Validation of an innovative open ended multimedia test. *European Journal of Work and Organizational Psychology*, *19*, 532–550.
- Oostrom, J., Lehmann-Willenbrock, N., & Klehe, U.-C. (2019). A new scoring procedure in assessment centers: Insights from interaction analysis. *Personnel Assessment and Decisions*, *5*, 73–82.
- Pincus, A. L., & Ansell, E. B. (2013). Interpersonal theory of personality. In H. Tennen, J. Suls, & I. B. Weiner (Eds.), *Handbook of psychology: Personality and social psychology* (p. 141–159). John Wiley & Sons, Inc..
- Pinsight. (2019, June). *Shorter & faster: Recent trend in Assessment Centers*. <https://www.pinsight.com/blog/2019/3/5/recent-trend-in-assessment-centers>.
- Ployhart, R. E., & Bliese, P. D. (2006). Individual adaptability (I-ADAPT) theory: Conceptualizing the antecedents, consequences, and measurement of individual differences in adaptability. In S. Burke, L. Pierce, & Salas (Eds.), *Understanding adaptability: A prerequisite for effective performance within complex environments* (pp. 3–39). San Diego: Elsevier.
- Ployhart, R. E., Lim, B. C., & Chan, K. Y. (2001). Exploring relations between typical and maximum performance ratings and the five factor model of personality. *Personnel Psychology*, *54*, 809–843.
- Ployhart, R. E., Schmitt, N., & Tippins, N. T. (2017). Solving the supreme problem: 100 years of selection and recruitment at the Journal of Applied Psychology. *Journal of Applied Psychology*, *102*, 291–304.

- Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology, 85*, 612–624.
- Purkiss, S. L., Segrest, Perrewe, P. L., Gillespie, T. L., Mayes, B. T. & Ferris, G. R. (2006). Implicit sources of bias in employment interview judgments and decisions. *Organizational Behavior and Human Decision Processes, 101*, 152–167.
- Putka, D. J., Le, H., McCloy, R. A., & Diaz, T. (2008). Ill-structured measurement designs in organizational research: Implications for estimating interrater reliability. *Journal of Applied Psychology, 93*, 959–981.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879–891.
- R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rauthmann, J. F., & Sherman, R. A. (2020). The situation of situation research: Knowns and unknowns. *Current Directions in Psychological Science, 29*, 473–480.
- Revelle W (2020). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois.
- Rivera, L. A. (2012). Hiring as cultural matching: The case of elite professional service firms. *American Sociological Review, 77*, 999–1022.
- Rivera, L. A. (2015). Go with your gut: Emotion and evaluation in job interviews. *American Journal of Sociology, 120*, 1339–1389.
- Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology, 85*, 370–395.
- Sadler, P., Ethier, N., Gunn, G. R., Duong, D., & Woody, E. (2009). Are we on the same wavelength? Interpersonal complementarity as shared cyclical patterns during interactions. *Journal of Personality and Social Psychology, 97*, 1005–1020.
- Sadler, P., Ethier, N., & Woody, E. (2011). Interpersonal complementarity. *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions* (pp. 123–142). New York: Wiley.
- Sadler, P., & Woody, E. (2016). *Manual for the Continuous Assessment of Interpersonal Dynamics (CAID) Joystick Monitor Program*.
- Sadler, P., Woody, E., McDonald, K., Lizdek, I., & Little, J. (2015). A lot can happen in a few minutes: Examining dynamic patterns within an interaction to illuminate the interpersonal nature of personality disorders. *Journal of Personality Disorders, 29*, 526–546.
- Short, J.A., Williams, E. and Christie, B. (1976) *The social psychology of telecommunication*. London: Wiley.
- Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology, 72*, 146–148.
- Strong, S. R., Hills, H. I., Kilmartin, C. T., DeVries, H., Lanier, K., Nelson, B. N., ... Meyer III, C. W. (1988). The dynamic relations among interpersonal behaviors: A test of complementarity and anticomplementarity. *Journal of Personality and Social Psychology, 54*, 798–810.
- Swider, B. W., Barrick, M. R., & Harris, T. B. (2016). Initial impressions: What they are, what they are not, and how they influence structured interview outcomes. *Journal of Applied Psychology, 101*, 625–638.

- Swider, B. W., Harris, T. B., & Gong, Q. (2022). First impression effects in organizational psychology. *Journal of Applied Psychology, 107*, 346–369
- Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. L. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology, 90*, 692–709.
- Thomas, K. M., Hopwood, C. J., Woody, E., Ethier, N., & Sadler, P. (2014). Momentary assessment of interpersonal process in psychotherapy. *Journal of Counseling Psychology, 61*, 1–14.
- Tiedens, L. Z., & Fragale, A. R. (2003). Power moves: Complementarity in dominant and submissive nonverbal behavior. *Journal of Personality and Social Psychology, 84*, 558–568.
- Tiedens, L. Z., Unzueta, M. M., & Young, M. J. (2007). An unconscious desire for hierarchy? The motivated perception of dominance complementarity in task partners. *Journal of Personality and Social Psychology, 93*, 402–414.
- Tracey, T. J. G. (2004). Levels of interpersonal complementarity: A simplex representation. *Personality and Social Psychology Bulletin, 30*, 1211–1225.
- Tracey, T. J. G., Bludworth, J., & Glidden-Tracey, C. E. (2012). Are there parallel processes in psychotherapy supervision? An empirical examination. *Psychotherapy, 49*, 330–343.
- Tracey, T. J. G., Ryan, J. M., & Jaschik-Herman, B. (2001). Complementarity of interpersonal circumplex traits. *Personality and Social Psychology Bulletin, 27*, 786–797.
- Tjosvold, D., Wong, A. S. H., & Chen, N. Y. F. (2014). Constructively managing conflicts in organizations. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 545–568.
- Van Iddekinge, C. H., Raymark, P. H., Roth, P. L., & Payne, H. S. (2006). Comparing the psychometric characteristics of ratings of face-to-face and videotaped structured interviews. *International Journal of Selection and Assessment, 14*, 347–359.
- Vowels, M. J., Vowels, L. M., & Wood, N. D. (2021). Spectral and cross-spectral analysis—A tutorial for psychologists and social scientists. *Psychological Methods. Advance online publication.*
- Waldman, D. A., Atwater, L. E., & Davidson, R. A. (2004). The role of individualism and the five-factor model in the prediction of performance in a leaderless group discussion. *Journal of Personality, 72*, 1-28.
- Wiggins, J. S. (1980). Circumplex models of interpersonal behavior. In L. Wheeler (Ed.), *Review of personality and social psychology* (pp. 265–294). Beverly Hills, CA: SAGE.
- Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management, 17*, 601–617.
- Wilson, K. S., DeRue, D. S., Matta, F. K., Howe, M., & Conlon, D. E. (2016). Personality similarity in negotiations: Testing the dyadic effects of similarity in interpersonal traits and the use of emotional displays on negotiation outcomes. *Journal of Applied Psychology, 101*, 1405–1421.
- Yao, Q., & Moskowitz, D. S. (2015). Trait agreeableness and social status moderate behavioral responsiveness to communal behavior. *Journal of Personality, 83*, 191–201.
- Yukl, G. (1999). An evaluative essay on current conceptions of effective leadership. *European Journal of Work and Organizational Psychology, 8*, 33–48.
- Zhou, L., Wang, M., & Zhang, Z. (2021). Intensive longitudinal data analyses with dynamic structural equation modeling. *Organizational Research Methods, 24*, 219-250.

**Table 1**

*Overview of Content and Interpersonal Demands of Role-plays*

	Low affiliation	High affiliation
High dominance	<p><i>Role-play 2:</i> Role-player criticizes candidate for slow decision-making.</p>	<p><i>Role-play 4:</i> Friendly role-player proposes to change the core event activity and seeks a solution for all stakeholders.</p>
Low dominance	<p><i>Role-play 1:</i> Role-player mentions a double booking but is not motivated to come up with a solution.</p>	<p><i>Role-play 3:</i> Friendly role-player asks for advice to maximize the event earnings and cooperates effectively.</p>

**Table 2**

*Interpersonal Momentary Adaptations in Affiliation and Dominance Between Candidates and Role-Players per Role-play*

Role-play	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Affiliation</i>			<i>Dominance</i>				
				% positive	<i>t(df)</i>	<i>p</i>	<i>M</i>	<i>SD</i>	% negative	<i>t(df)</i>	<i>p</i>
1	96	.24	.34	78	7.02(95)	< .001	-.48	.43	88	-10.89(95)	< .001
2	95	.22	.33	77	6.46(94)	< .001	-.50	.40	86	-12.19(94)	< .001
3	94	.23	.27	83	8.48(93)	< .001	-.64	.39	92	-16.04(93)	< .001
4	95	.30	.31	86	9.38(94)	< .001	-.71	.45	90	-15.30(94)	< .001

*Note.* *N* = 96. *M* refers to the mean within-person correlation. For affiliation, we expect a positive correlation because it indicates that as one party expresses high (low) affiliative behavior, the other party also responds with high (low) affiliative behavior. For dominance, a negative correlation is expected because it indicates that as one party expresses high dominant behavior, the other party responds with the opposite low dominant behavior and vice versa.

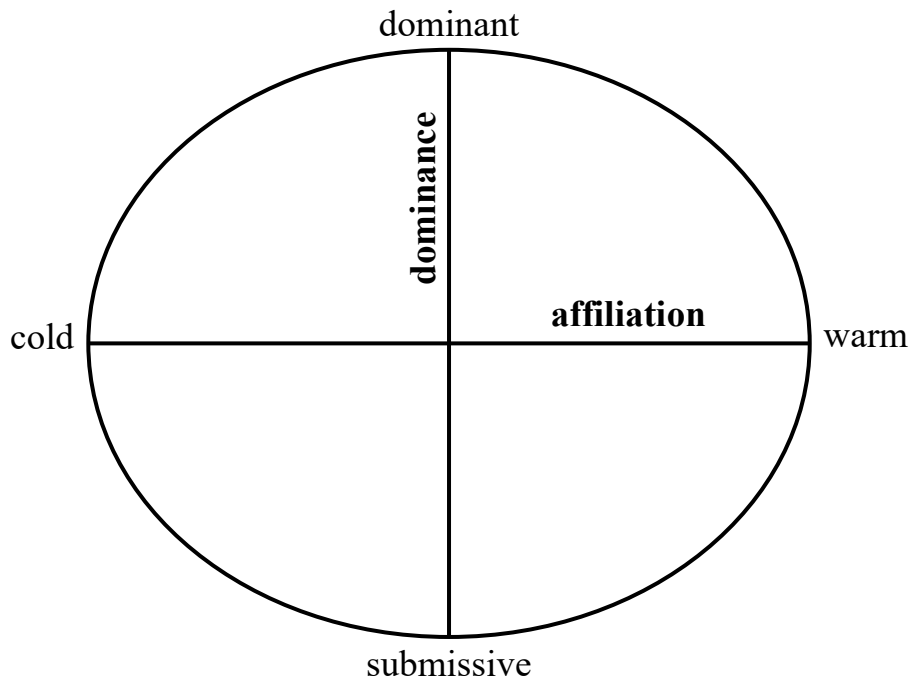
**Table 3**  
*Means, Standard Deviations, and Intercorrelations Between Study Variables*

Variable	<i>n</i>	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9.
<b>Demographics</b>												
1. Gender	96	1.51	0.50									
2. Age	96	23.62	1.85	-.15								
<b>Complementarity Variables</b>												
3. Affiliation influence	96	-0.99	0.43	-.12	-.03							
4. Dominance influence	96	1.33	0.61	.08	.03	-.24*						
5. Affiliation adaptation	96	0.25	0.16	-.09	-.07	.39**	-.14					
6. Dominance adaptation	96	-0.58	0.27	-.17	.11	-.15	.56**	-.12				
<b>Role-play Ratings</b>												
7. From <i>in-situ</i> assessors	96	5.46	1.08	-.07	-.28**	.21*	-.22*	.21*	-.26*			
8. From <i>ex-situ</i> assessors	96	5.62	0.80	.02	-.38**	.14	-.04	.13	-.14	.65**		
<b>Criterion Performance</b>												
9. Interpersonal adaptability	95	56.48	22.38	-.27**	-.11	.23*	-.06	.11	-.07	.31**	.33**	
10. Task performance	95	60.17	22.46	-.08	-.34**	.00	-.13	-.16	-.22*	.34**	.44**	.50**

*Note.* *N* = 96. Gender: male = 1, female = 2. Higher values represent higher/lower complementarity for affiliation/dominance. \* *p* < .05, \*\* *p* < .01

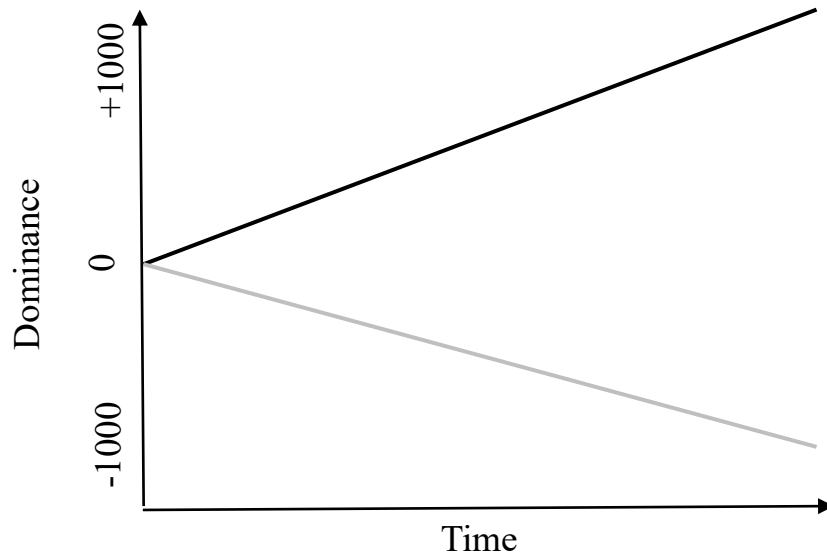
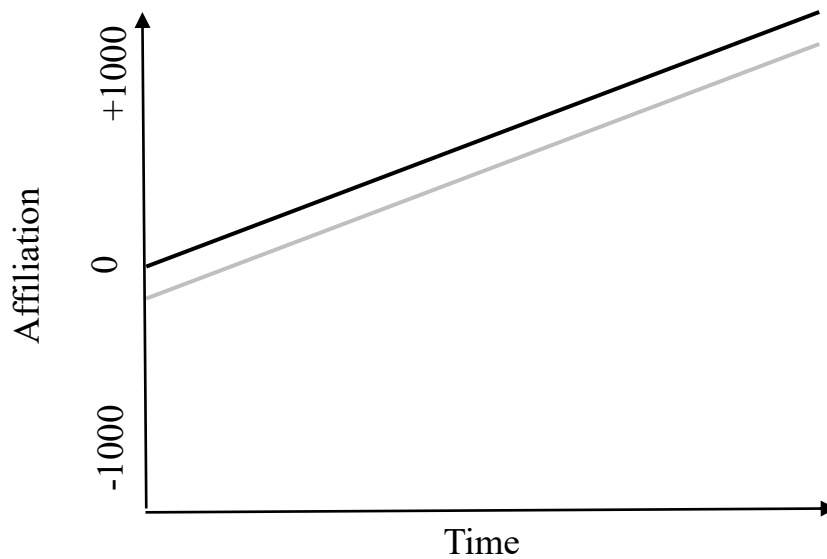


**Figure 1**  
*The Interpersonal Circumplex*



**Figure 2**

*Example of Mutual Influence Trend for Dominance (Figure 2a) and Affiliation (Figure 2b)*

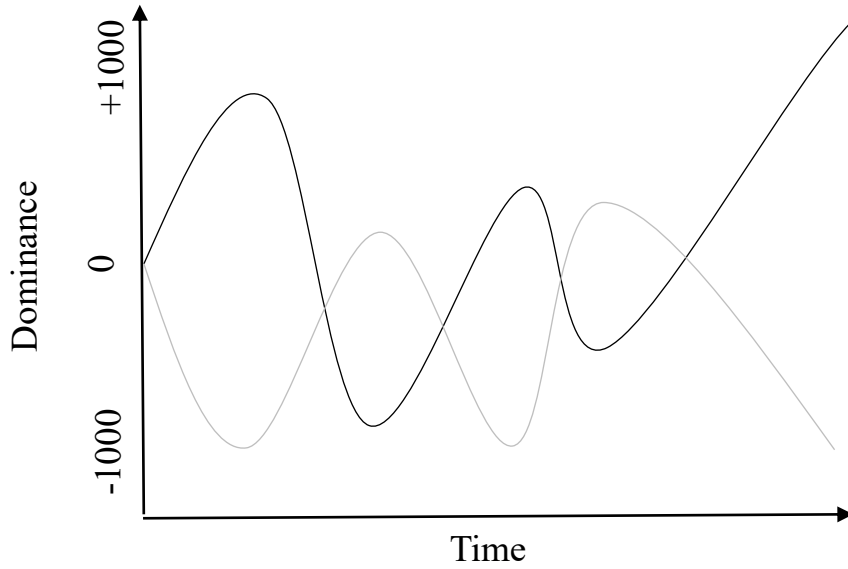
**Figure 2a****Figure 2b**

*Note.* The black and grey lines refer to interpersonal behavior of members of a dyad (candidate and role-player). Figure 2a exemplifies influence in dominance because the temporal dominance trends of the interaction partners run opposite to each other. Figure 2b exemplifies influence in affiliation because the temporal affiliation trends of the interaction partners are entrained.

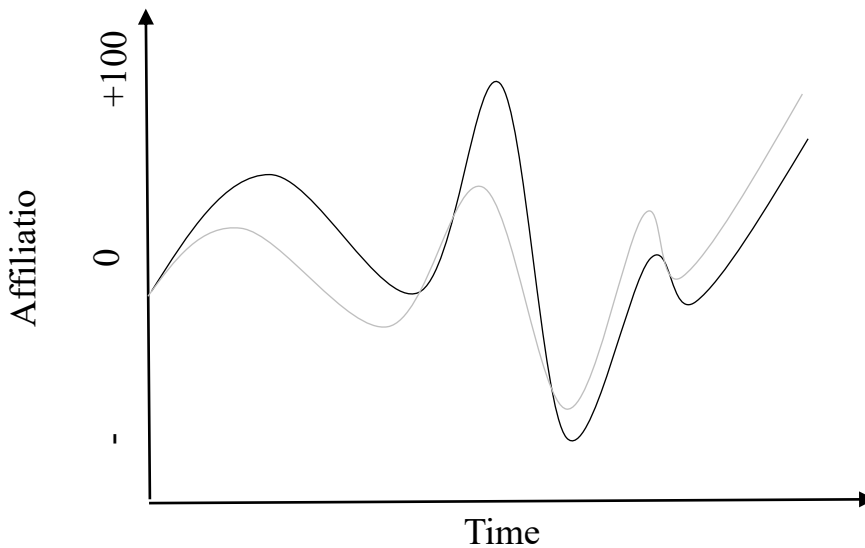
**Figure 3**

*Example of Mutual Adaptations for Dominance (Figure 3a) and Affiliation (Figure 3b)*

**Figure 3a**



**Figure 3b**

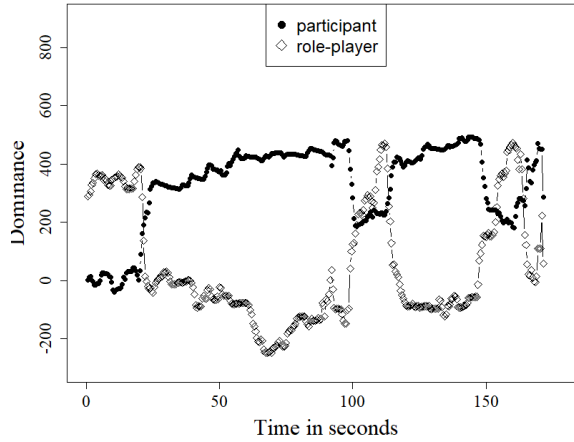


*Note.* The black and grey lines refer to interpersonal behavior of members of a dyad (candidate and role-player). Figure 3a exemplifies mutual adaptations in dominance because expressions of dominant behavior by one interaction partner are mirrored by expressions of following behaviors by the other one. Figure 3b exemplifies mutual adaptations in affiliation because expressions of affiliative behavior by one interaction partner are mirrored by affiliative expressions by the other one.

**Figure 4**

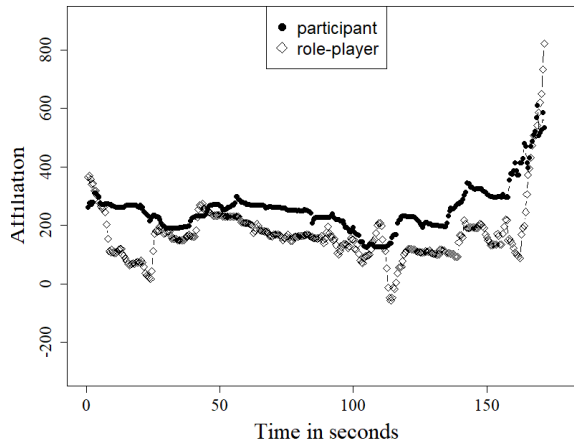
*Example Bivariate Time Series Related to a Dyad’s Momentary Levels Across the Role Play*

**Figure 4a: Dominance**



*Note.* Cross-correlation:  $r = -.84, p < .001$

**Figure 4b: Affiliation**



*Note.* Cross-correlation:  $r = .67, p < .001$

*Note.* The black and grey lines refer to interpersonal behavior of members of a dyad (candidate and role-player). Figure 4a displays mutual adaptations in dominance in one example dyad as expressions of dominant behavior by one interaction partner are mirrored by expressions of following behaviors by the other one. Figure 4b shows mutual adaptations in affiliation in one example dyad as expressions of affiliative behavior by one interaction partner are mirrored by expressions of affiliative behavior by the other one.

### Appendix

We conducted a pilot study to examine whether the four role-plays indeed reflected the different interpersonal demands. Two female graduate students in I/O psychology with expertise in the interpersonal circumplex rated the role-players' overall expression of affiliation and dominance as depicted in the role-player instructions. They provided their ratings via the interpersonal grid (Moskowitz & Zuroff, 2005), which projects the interpersonal circumplex to a Cartesian plane consisting of 11x11 boxes with adjectives clarifying the different quadrants of the interpersonal circumplex. Raters then ticked a box that best matched the interpersonal behavior of a target person, leading to a score for affiliation and dominance that can vary between 1 and 11. Ratings of the role-player instructions showed moderate to excellent interrater reliabilities ( $ICC[2,1] = .96$  for affiliation;  $ICC[2,1] = .65$  for dominance) and generally confirmed the purported overall interpersonal behavior for role-players. Across raters, both role-plays with relatively low prescribed affiliation received affiliation ratings of 3.5, whereas role-plays with relatively high prescribed affiliation received affiliation ratings of 8.50 (role-play 3) and 9.00 (role-play 4). Role-plays with relatively low prescribed dominance received dominance ratings of 5.5 (role-play 1) and 6 (role-play 3), whereas role-plays with prescribed relatively high dominance received dominance ratings of 10 (role-play 2) and 7.5 (role-play 4).

In addition, we used the CAID codings to investigate role-players' *average* interpersonal behavior over time series per role-play. Results showed that role-players' behavior varied across the affiliation dimension: As expected, role-players overall showed low values for affiliation in role-play 1 ( $M = -47.20$ ,  $SD = 104.07$ ,  $95\%CI = [-68.29; -26.12]$ ) and role-play 2 ( $M = -120.52$ ,  $SD = 147.81$ ,  $95\%CI = [-150.63; -90.41]$ ), whereas the values were higher in role-play 3 ( $M = 183.10$ ,  $SD = 98.05$ ,  $95\%CI = [163.02; 203.18]$ ) and role-play 4 ( $M = 93.73$ ,  $SD = 95.27$ ,  $95\%CI = [74.33; 113.14]$ ). Role-players' overall behavior also varied across the dominance continuum. As expected, the highest values for expressing dominance were found for role-plays 2 ( $M = 236.68$ ,  $SD = 104.51$ ,  $95\%CI = [215.39; 257.97]$ ) and 4 ( $M = 162.54$ ,  $SD = 117.05$ ,  $95\%CI = [138.69; 186.38]$ , respectively), whereas role-plays 1 and 3 had the lower values ( $M = 47.21$ ,  $SD = 129.00$ ,  $95\%CI = [21.07; 73.34]$  and  $M = 146.06$ ,  $SD = 91.69$ ,  $95\%CI = [127.28; 164.85]$ , respectively).