This is a pre-copy edited, author-produced version of an article accepted for publication in *Leadership Quarterly* following peer review.

The version of record is available online at: 10.1016/j.leaqua.2020.101473

Citation:

Klonek, F. E., Gerpott, F. H., & **Parker, S. K.** (2020). A conceptual replication of ambidextrous leadership theory: An experimental approach. *The Leadership Quarterly,* 101473. https://doi.org/10.1016/j.leaqua.2020.101473 (ABDC **A*** journal).

A Conceptual Replication of Ambidextrous Leadership Theory: An Experimental Approach

Innovation—the introduction, promotion, and implementation of new ideas (Hughes, Lee, Tian, Newman, & Legood, 2018; West & Farr, 1990)—is pivotal for building and maintaining organizational competitiveness (Artz, Norman, Hatfield, & Cardinal, 2010; Rubera & Kirca, 2012; Tushman & O'Reilly, 1997). Whereas a significant amount of research emphasizes the importance of leadership for motivating followers to achieve innovation success (e.g., Denti & Hemlin, 2012; Mumford, Scott, Gaddis, & Strange, 2002; Hughes et al., 2018; Rosing, Frese, & Bausch, 2011), there is an ongoing scholarly debate about the specific set of leader behaviors that is suitable in this process (Mumford et al., 2002; Stock, Zacharias, & Schnellbaecher, 2016). This discussion is based on the notion that traditionally studied leadership styles, such as transformational leadership, are too unspecific to account for the complex requirements of innovation work (Bledow, Frese, Anderson, Erez, & Farr, 2009; Hughes et al., 2018; Rosing, Rosenbusch, & Frese, 2010). Particularly, scholars argue that leaders need to engage in behaviors that align with the dynamic nature of the innovation process, which involves both the exploratory search of creative ideas, and the exploitative selection and implementation of ideas (Bledow et al., 2009; Rosing et al., 2011).

To address this call for studying leadership behaviors that are theoretically relevant to the innovation process, a recent research stream has brought forward the concept of ambidextrous leadership, defined as the complementary engagement in two distinct types of leadership activities (Rosing et al., 2011; for empirical articles testing this theory see Table 1). Specifically, ambidextrous leaders switch between opening behaviors (i.e., aimed at enhancing variability in follower behaviors) and closing behaviors (i.e., seeking to reduce variability in follower behaviors)

in consideration of the continuously changing task demands (Rosing et al., 2011). On the one hand, leader opening behaviors allow employees to make errors, encourage them to use alternative methods to accomplish their tasks, and motivate them to take risks. That is, opening behaviors stimulate employees' exploratory variance-increasing "search" behaviors (March, 1991; Rosing et al., 2011). On the other hand, leader closing behaviors focus on establishing work routines, goal monitoring, and rule adherence. That is, closing behaviors encourage employees' variance-decreasing "production" behaviors (March, 1991, Rosing et al., 2011). By switching between opening and closing behaviors according to the task progress, leader ambidexterity (i.e., the interaction between leader opening and closing behaviors) is assumed to drive follower ambidexterity (i.e., the interaction between follower exploration and exploitation behaviors). This interplay ultimately improves innovation outcomes because innovation requires the development (i.e., exploration) and the implementation (i.e., exploitation) of ideas in equal means (March, 1991).

An ambidextrous view on leadership for innovation has attracted considerable interest. To illustrate, Google Scholar reports 576 and the Web of Science 172 citations[1] of the work of Rosing et al. (2011) that developed ambidextrous leadership theory after meta-analytically finding that a range of leadership styles (particularly, transformational and transactional leadership) showed heterogeneous associations with innovation. Practitioner journals have also enthusiastically picked up the topic with articles such as "The ambidextrous CEO" (Tushman, Smith, & Binns, 2011) or "How to become an ambidextrous leader" (Kinni, 2016). Universities even offer programs that claim to train managers to increase their ambidexterity level (e.g., Certificate in Innovative Leadership, David Eccles School of Business, 2017). Yet, despite the

[1] Retrieved March 2019

widespread interest, the hypothesized relationships of ambidextrous leadership theory have not been tested by designs that allow causal conclusions. As we discuss in more detail below, scholars have thus far relied on survey-based correlational designs and acknowledged that this limits causal interpretation of their results (e.g., Alghamdi, 2018; Zacher et al., 2016; Zacher & Rosing, 2015). Accordingly, these researchers have also suggested to use controlled experimental approaches to test the suggested relationships of leadership ambidexterity theory. We follow this call by proposing a conceptual replication of ambidextrous leadership theory across two studies. This means that we purposefully adapt previously used methods to test the theory's underlying assumptions (Makel, Plucker, & Hegarty, 2012).

Our research provides two main contributions to the literature. First, from a methodological perspective, we seek to offer a stronger empirical foundation for the necessity of an innovation-specific leadership style, namely ambidextrous leadership. Specifically, the present research uses experimental research designs with external ratings of innovation and an instrumental variable approach to conceptually replicate the results from previous studies.

Second, on a more general level, we assume that our replication endeavor is important for an advancement of the research field. If a theory receives considerable scholarly attention—"as an arbitrary selection, if a publication is cited 100 times" (Makel et al., 2012: 541)—a replication of findings beyond the original authors of that theory can help to stimulate research designs that have stronger confirmatory power (Schmidt, 2009) and that test the underlying relationships in a controlled environment. This is also important from a practitioner perspective. The popularization of any leadership theory in mainstream media can strongly influence managerial practice, which is problematic in case the existing evidence for a theory suffers from methodological artifacts, cannot be generalized to different populations, or does not reflect true causal relationships. As others have warned, practitioners tend to rely "on popular ideas and fads without sufficient consideration given

to the validity of these ideas" (Zaccaro & Horn, 2003: 779). In the worst case, managers are trained on a leadership style that may be useless or could potentially harm innovation outcomes. We hope that our research helps to provide stronger causality of the previous findings on leadership and innovation, and thus ultimately contributes to building a more integrated and practically relevant leadership field.

In what follows, we first describe the assumptions of ambidextrous leadership theory in more detail and explain the rationale of the hypotheses (Figure 1). To establish why a conceptual replication is needed, we then review methodological shortcomings of existing studies. We subsequently describe the design, measures, and analytical strategies of the present research.

++++++ Insert Figure 1 about here++++++

Theoretical Model of Ambidextrous Leadership for Innovation

Scholars have long sought to understand how organizations can foster employee innovation by studying organizational-level predictors (e.g., Camisón-Zornoza, Lapiedra-Alcamí, Segarra-Cipés, & Boronat-Navarro, 2004) and team-level predictors (e.g., Hülsheger, Anderson, & Salgado, 2009). Particularly, innovation scholars have put forward the idea that the development of innovations involves a set of complementary variance-increasing activities (i.e., exploration of ideas) and variance-decreasing activities (i.e., selection and implementation of ideas) that align with different phases of the innovation process (Hughes et al., 2018; West & Farr, 1990). Recent research has transferred this idea to the interpersonal level. Acknowledging the role of leadership in this process, scholars have suggested that the innovation process can be managed effectively by a set of complementary ambidextrous leadership behaviors (e.g., Rosing et al., 2011).

Specifically, the development of ambidextrous leadership was motivated by a meta-analysis summarizing associations between innovation and different leadership styles, particularly transformational and transactional leadership (Rosing et al., 2011). In their study, the

authors found that the links between innovation and transformational as well transactional leadership were moderate, and varied largely between studies (see Rosing et al., 2011). From this evidence, Rosing et al. (2011) concluded that the existing constructs do not sufficiently capture the leader's focus on increasing and reducing variance in followers' behavior, two aspects that are essential in the innovation process (West & Farr, 1990). Theoretically, adding such a focus should result in a more aligned leadership style that captures the dynamics of innovation tasks, such that "a leadership style is positively related to innovation when complemented by another leadership style that focuses on and fosters different aspects of the innovation process" (Rosing et al., 2011: 965). Notably, while other leadership styles such as transformational or transactional leadership can also co-occur, they may or may not include the increase or reduction of variance in follower behaviors. As such, these leadership theories do not capture the central element of variability, which is essential for the definition of ambidextrous leadership.

According to the work by Rosing and colleagues (2011), ambidextrous leadership consists of opening behaviors (i.e., activities increasing variance in followers, such as encouraging experimentation and attempts to challenge established approaches or giving room for independent thinking) and closing behaviors (i.e., activities decreasing variance in followers, such as monitoring goal achievement, taking corrective action, or setting specific guidelines)[2]. It is noteworthy that the concept of leadership ambidexterity (i.e., a set of complementary behaviors that trigger respective complementary follower behaviors) is defined by its outcomes. Thus, the

[2] Providing support for the conceptual distinctiveness of ambidextrous leadership, previous research could clearly distinct leader opening behaviors from transformational leadership and leader closing behaviors from transactional leadership. Specifically, although opening behaviors and transformational leadership are positively correlated (e.g., *r* = .67** in Zacher et al., 2016; *r* = .49** in Zacher & Rosing, 2015), confirmatory factor analyses have shown that they do not reflect the same underlying construct. Similarly, closing behaviors and transactional behaviors are positively correlated (e.g., r = .48** in Zacher et al., 2016; r = .48** in Zacher & Rosing, 2015), but can be clearly distinguished on a factor-analytical level.

clarity of the concept can be criticized by not covering what that the nature of leadership ambidexterity actually is (MacKenzie, 2003). In other words, leader opening behaviors are *defined* by their consequences (i.e., follower exploration behaviors, e.g., activities such as deviating from routine work, trying out new approaches, or expanding knowledge to fulfill the innovation task, Alghamdi, 2018; Rosing et al., 2011; Zacher et al., 2016) and they are also *hypothesized* to stimulate exploration behavior. In a similar way, leader closing behaviors are both conceptualized and theorized as antecedents of follower exploitation behaviors (e.g., engaging in standardized or routine activities, applying present work knowledge in the task at hand, or focusing on implementing well-defined tasks, Alghamdi, 2018; Rosing et al., 2011; Zacher et al., 2016). Closing behaviors thus reduce variance in follower behavior (and the theory argues that opening is often required in early whereas closing is required in later phases of innovation tasks when ideas need to be implemented). These assumptions on the positive association between leader opening and closing behaviors and follower exploration and exploitation behaviors are summarized in the following hypotheses:

H1: Opening leader behaviors positively predict follower explorative behaviors.

H2: Closing leader behaviors positively predict follower exploitative behaviors.

Two empirical studies have investigated these suggested positive associations (Alghamdi 2018, Zacher et al., 2016[3]). First, Zacher et al. (2016) used a cross-sectional survey design with 388 employees recruited via the online platform MTurk. Participants were asked to think of their leader and subsequently assess their leaders' opening and closing behaviors as well as their own exploration and exploitation behaviors at work. Results showed that leader opening behaviors

[3] We use the word "predict" in our hypotheses to specify causal relationships of the model. However, we acknowledge that previous studies have worded their hypotheses more cautiously and used the term "positively associated" for H1 and H2.

explained a significant amount of variation in follower exploration (supporting H1) and that leader closing behaviors explained a significant variation in follower exploitation (supporting H2). Notably, the authors controlled for transformational leadership, transactional leadership, and employee person-level factors (i.e., openness to experience, conscientiousness, positive trait affect). Second, Algahmdi (2018) used a single-source self-report study with 147 faculty members in Saudi Arabia. Controlling for gender, educational level, and faculty position, the author showed that supervisors' opening behaviors were positively associated with faculty member exploration (supporting H1) and supervisors' closing behaviors were positively associated with faculty member exploitation (supporting H2).

Leader ambidexterity and follower innovation

Opening and closing behaviors together form the leader ambidexterity construct. Ambidextrous leadership theory states that leaders need to engage in both behaviors in line with the innovation task to increase innovation outcomes. Unfortunately, the original theory did not precisely explain when or how these behaviors should be expressed but it proposed that the requirements of the innovation task may require different degrees of variability in follower behaviors (i.e., exploration and exploitation). Hence, if leaders encourage followers to show these behaviors at the right point in time, the outcome of the innovation task will be improved. Because these complementary behaviors (i.e., opening and closing) are expected to interactively shape innovation, the higher-order construct of ambidextrous leadership is traditionally captured as a multiplicative interaction of these sub-dimensions (Zacher & Wilden, 2014; see also Gibson & Birkinshaw, 2004 for measuring related constructs such as contextual ambidexterity; and Mom, Van Den Bosch, Volberda, 2009 for managerial ambidexterity). Stated formally, this translates into the following hypothesis:

H3: The interaction of leader opening and closing behaviors (i.e., leader ambidexterity) positively predicts employee innovation, insofar that innovation is highest when both leader opening and leader closing behaviors are high.

Three studies have tested the moderation effect of leader opening and closing behaviors on follower innovation (Alghamdi, 2018; Zacher & Rosing, 2015; Zacher & Wilden, 2014). First, a single-source self-report study by Algahmdi (2018) showed that the multiplicative interaction of leader opening and closing behaviors predicted faculty members' innovation performance above and beyond the main effects of opening and closing behaviors. Second, Zacher and Wilden (2014) conducted a diary study with 113 employees and found that followers' self-reported daily innovative work behavior was highest when followers perceived both leader opening and closing behaviors to be high on the same day (controlling for their respective main effects; leaders' daily intellectual stimulation, employees' trait positive affect, and employees' general level of job autonomy). Third, Zacher and Rosing (2015) conducted a team-level study by asking 33 team leaders to rate innovation performance of their teams and asking followers about their respective leaders' ambidextrous behaviors. The findings indicated a significant interaction effect for leader opening and closing behaviors on team innovation (controlling for transformational leadership and general team success). Specifically, leader opening behaviors were only related to team innovation when closing behaviors were also high.

Leader ambidexterity and follower behavior

In more recent models, scholars have extended the original model (e.g., Zacher et al. 2016; see also Rosing & Zacher, 2017) by suggesting that follower ambidexterity (i.e., the interaction of employee exploration and exploitation behaviors) is the more proximal antecedent of follower innovation. That is, instead of focusing on leader opening and closing behaviors and the direct link with employee innovation, these studies focus on the logical consequence of the

previously outlined hypotheses, namely that follower exploration and exploitation behaviors should be drivers of employee innovation. Stated formally:

> *H4: The interplay of follower exploration and exploitation (i.e., follower ambidexterity) positively predicts innovation outcomes, insofar that innovation is highest when both follower exploration and exploitation behaviors are high.*

The hypothesis that follower exploration and exploitation interactively shape follower innovation has also received support from Zacher et al. (2016) as well as Rosing and Zacher (2017). First, in their cross-sectional study with MTurk participants, Zacher et al. (2016) reported that self-reported innovation performance (rated by employees) was significantly predicted by self-rated exploration, exploitation, and their multiplicative interaction. The interaction indicated that employee exploitation was more strongly related to innovative performance when employees also showed high levels of exploration behavior. Second, Rosing and Zacher (2017) conducted two diary studies (one weekly and one daily study) that repeatedly asked employees to rate their weekly (Study 1) and daily (Study 2) levels of exploration behavior, exploitation behavior, and innovative work performance. The authors used polynomial regression with response surface analysis which is an alternative approach to examine the balance between the two conceptually related constructs of follower exploration and exploitation. The results of their studies showed that employees who showed high levels of both exploration and exploitation reported highest levels of innovative work performance (controlling for weekly levels of positive and negative affect, innovation requirements of the job, age, gender, and education; Rosing & Zacher, 2017). Furthermore, their study showed that an imbalance towards follower exploration was less harmful to innovation than an imbalance towards exploitation.

Table 1 summarizes the previous studies on ambidextrous leadership, their design, control variables, and main findings (see also Figure 1 for an illustration of the hypotheses and

its empirical support). Next, we summarize the main methodological shortcomings that limit our knowledge about the hypothesized relationships of ambidextrous leadership theory. These shortcomings comprise the starting point for our conceptual replication endeavor.

++++++ Insert Table 1 about here++++++

Lack of Rigorous Tests of Ambidextrous Leadership Theory

To test ambidextrous leadership theory, the majority of empirical studies have used single-source (Alghamdi, 2018; Rosing & Zacher, 2017; Zacher et al., 2016; Zacher & Wilden, 2014) and/or cross-sectional survey field studies (Alghamdi, 2018; Zacher et al., 2016; Zacher & Rosing, 2015). While field research has the advantage of providing a high ecological validity, a challenge is that many of these studies were conducted using cross-sectional designs. That is, they measure the independent and dependent variables concurrently, which entails that it is unclear whether ambidextrous leadership fosters follower ambidexterity or whether the extent to which followers engage in ambidextrous behaviors increases their leaders' probability to exhibit ambidextrous behaviors. Furthermore, participants' answer patterns may be driven by several uncaptured variable(s) which correlate with the modeled variables but are not included in the model. This describes an endogeneity problem due to common-method variance (Antonakis, Bendahan, Jacquart, & Lalive, 2010; Podsakoff, Mackenzie, & Podsakoff, 2012; Podsakoff & Organ, 1986), where the effect of a causal predictor (i.e., leader ambidexterity) and/or mediator (follower ambidexterity) on a dependent variable (innovation) cannot be interpreted because the predictor does not vary randomly and its effect on the dependent variable can be explained by other omitted co-variants (Antonakis et al., 2010). If ambidextrous leader or follower behaviors are endogenous regressors and the analytical methods do not address problems of endogeneity, then previous findings do not help to fully understand the phenomenon, or to put it more bluntly, "finding a relationship between an endogenous regressor x—that has not been purged from

endogeneity somehow—and y does not help leadership theory one bit" (Antonakis et al., 2014: 94).

In order to overcome endogeneity in the model, the independent variable and/or mediator must be exogenous, meaning it must be unaffected by any other variable in the model (Antonakis et al., 2014). That is, there must be no variable that would be an antecedent of innovation and would correlate with ambidexterity (i.e., no omitted variable). Yet, there are reasons to assume that ambidexterity is predicted by certain unmodeled causes that directly affect employee ambidexterity and/or innovation performance.[4] For instance, empirical evidence indicates that conscientiousness is negatively associated with ambidexterity (Keller & Weibler, 2015); yet, conscientiousness is also a direct antecedent of innovation success (Stock, Van Hippel, & Gillert, 2016).

In addition to endogeneity concerns, the exclusive use of survey designs in leadership ambidexterity research can be criticized for reasons of reactivity (i.e., subjects change their responses because they are sensitized for the construct under investigation; Hill, White, & Wallace, 2014), and for self-serving and retrospective biases (Baumeister, Vohs, & Funder, 2007). In other words, perceptions of behavior can differ considerably from actual objective behavior (Behrendt, Matz, & Göritz et al., 2016). Notably, previous research is well aware of these challenges, and has acknowledged these limitations of self-reports in the future research section (e.g., Zacher et al., 2016; Zacher & Wilden, 2015; Rosing & Zacher, 2017).

---

[4] Previous studies have included multiple control variables that are conceptually argued to predict both ambidextrous leadership and innovation outcomes (for an overview of all control variables, see Table 1). However, a simple estimating system of equations whereby ambidextrous leadership is modeled as an outcome of these controls does not produce the correct estimates if ambidextrous leadership is endogenous and no two-stages-least-squares estimation is used (Antonakis et al., 2010). In other words, including these constructs as control variables does not solve the endogeneity problem because it does not instrumentalize the ambidextrous leadership construct.

To summarize, based on previous research that has provided initial support for ambidextrous leadership theory in the field, we consider the rime ripe to conduct a conceptual replication that can help address the limitations from previous survey-based and correlational research designs. As others have argued, once empirical data from non-experimental field research "suggest a moderation effect might be present, investigators must devote the additional time, attention, and resources needed to perform more rigorous experimental (…) designs to reveal the true latent model" (Murhpy & Russel, 2017: 558). Hence, we decided to replicate the results from leadership ambidexterity theory by using an experimental design that has significant "confirmatory power" to corroborate that ambidextrous leadership also works in principle and "in isolation" (without all the other factors present in field environments). By doing this, our approach is a conceptual replication (which is defined as a repetition of a test of a hypothesis of earlier research work with different methods, that is, by using a different material realization, i.e., research design, Schmidt, 2009). In contrast to a direct (or close) *replication* – which is the most exact possible duplication of previous research procedures (ideally even in the same lab; Brandt et al., 2014; Schmidt, 2009) and which mainly serves to (re)-produce scientific facts – the main purpose of a conceptual replication is to verify the underlying relationships hypothesized in a model (i.e., to extend knowledge and produce understanding, Schmidt, 2009). Given methodological challenges that often characterize the original studies, conceptual replications therefore regularly need to use a different material realization (i.e., research design, Schmidt, 2009).

Study 1: Randomized Vignette Experiment

Study 1 replicates previous research (i.e., Alghamdi, 2018; Zacher et al., 2016; Rosing & Zacher, 2017; Zacher & Rosing, 2015; Zacher & Wilden, 2014) by testing the causal effects of

leader ambidexterity on follower ambidexterity and innovative performance rated by subject matter experts through a randomized experimental study with four conditions (i.e., opening, closing, ambidextrous, and a control condition of transformational leadership). An experimental vignette methodology provides a systematic approach to ensure high internal validity (Aguinis & Bradley, 2014). Particularly, we aim to test in Study 1 whether leader opening behaviors enhances follower exploration (H1) and whether leader closing behaviors enhance follower exploitation (H2, both hypotheses are conceptual replications of Alghamdi, 2018; Zacher et al., 2016). Furthermore, we investigate the interactive effect of leader opening and closing (i.e., ambidexterity) on innovation (H3, replications of Alghamdi, 2018; Zacher & Rosing, 2015; Zacher & Wilden, 2014) as well as the interactive effect of follower ambidexterity on innovation (H4, conceptual replication of Rosing & Zacher, 2017; Zacher et al., 2016).

Methods

The study is subject to approval by the institutional review board of the XX University prior to conducting the study. Study data will be published in anonymized form on the Open Science Framework.

Sample. We used two complementary approaches to determine sample size using an a priori power analyses: First, we reviewed effect sizes reported in the literature and used G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) to calculate required sample size (for details see Appendix A for a full protocol). Second, we also reviewed the sample sizes of previous research (see Appendix B) and compared them with the results of our estimations.

For H1, reported effects have been r =.41 to .42 for the link between opening and exploration (i.e., H1, Alghamdi, 2018, Zacher et al., 2016). For H2, effect sizes have been r =.45 to .21 for the link between closing and exploitation (Alghamdi, 2018, Zacher et al., 2016). For

H3, effect sizes have been $f^2 = .06$ (Algahmdi, 2018) to $f^2 = .21$ (Zacher & Rosing, 2015) for the link between leader ambidexterity and innovation. For H4, effect sizes have been $f^2 = .01$ for the link between follower ambidexterity and innovation (Zacher et al., 2016)[5].

We calculated necessary sample sizes for each of the four hypotheses separately (using $\alpha = .05$ and a high power of $\beta = .95$ based on recommendations in the replication literature, Open Science Collaboration, 2012). Furthermore, when more than one reported effect size was available, we relied on the smaller value (e.g., we used $r = .21$, not $r = .45$ as an effect size measure for H2) to have the most conservative estimate. For H1, the necessary sample size to detect an effect size of $r = .41$ is $N = 71$. For H2, the necessary sample size to detect an effect size of $r = .21$ is $N = 289$. For H3, the necessary sample size to detect a moderation effect $f^2 = .05$ is $N = 262$. For H4, the necessary sample size to detect a moderation effect $f^2 = .01$ is $N = 360$. We considered that some of the participants may complete the task insufficiently (e.g. not passing attention check items). Hence, we also considered an over-recruitment of 110%, resulting in a required sample of N = 396 persons for the experiment (hence $n = 100$ per condition).

Furthermore, we used an alternative suggested approach for determining sample sizes for replication research (Brandt et al., 2014). This approach suggests taking 2.5 times the sample size reported in the original study (Simonsohn, 2013). To obtain this estimate, we calculated the average sample size for those studies that have tested previous hypotheses in the LAT model (i.e., all studies in Table 1). The average sample from these studies is N = 157 (see Appendix B

[5] We could not calculate a local effect size f^2 based on the information provided in Rosing and Zacher (2017) and Zacher and Wilden (2014). These studies have used more complex hierarchical data structures (i.e., diary studies with observations nested within subjects) which do not allow a straightforward estimation of effect sizes.

for details); hence a 2.5 times this estimate would result in sample size of N = 392.5 which is almost identical to our planned sample size of $N = 400$.

Recruitment. We will collect data from working professionals recruited through TurkPrime (Litman et al., 2016) a versatile crowdsourcing data acquisition platform for researchers that allows us to recruit participants using pre-screening requirements (Litman, Robinson, & Abberbock, 2017; Keith, Tay, & Harns, 2017). Research comparing the validity of results from participants tested via different channels (i.e., face-to-face laboratory settings, social media posts, and platforms like TurkPrime or MTurk) indicates that crowdsourcing data reach equivalent—sometimes even superior—data quality than data collected in in-person settings, if researchers make careful decisions about the design and applied analytical procedures (Casler, Bickel, & Hackett, 2013, Cheung, Burns, Sinclair, & Sliter, 2017; Thomas & Clifford, 2017). Therefore, our study will only be accessible for participants who have an approval rate of ≥ 90% and we will include multiple measures of attention/ comprehension checks which ensure high data quality (i.e., screener items, cf., Thomas & Clifford, 2017).

To allow for better comparisons with populations from previous research, we carefully reviewed the sample information (i.e., see Table 1 and Appendix B) and based our pre-selection criteria on this information. In terms of recruitment, the majority of previous research has conducted online surveys (i.e., Rosing & Zacher, 2017; Zacher et al., 2016; Zacher & Wilden, 2016) and one study used a sample of Mturk participants (Zacher et al., 2016). In terms of nationality, previous research used samples from Australia (Zacher & Rosing, 2015; Rosing & Zacher, 2017), USA (Zacher et al., 2016), Germany (Rosing & Zacher), and Saudi-Arabia (Alghamdi, 2018). In terms of industries, previous research has surveyed participants working in public universities (Alghamdi, 2018), creative industries (Rosing & Zacher, 2017), architecture

and design firms (Zacher & Rosing, 2015), or from a "range of industries" (Rosing & Zacher, 2017, including participants working in non-creative industries like administration officers and even cleaners, Zacher & Wilden, 2014).

Hence, our study is only accessible for participants who work full-time (≥ 35+ hours), are located in Australia or the US (to ensure language abilities and to match population criteria with Zacher & Rosing, 2015; Rosing & Zacher, 2017), are at least 18 years of age, and have a direct line manager/supervisor (to match selection criteria with Zacher et al., 2016).

Participants will be paid USD 9.00 for their participation (remuneration is slightly higher than the US minimum wage of USD 7.25/hour).

Experimental task.

The experimental task is designed in a way so that it captures a proxy of the workplace innovation from previous research (e.g., Zacher et al., 2016; Zacher & Rosing, 2015). In this previous research on leader ambidexterity, employee innovation was rated using four items that assessed the extent to which employees showed one of the following behaviors: "coming up with new ideas", "working to implement new ideas," "finding improved ways to do things," and "creating better processes and routines." (e.g., Zacher et al., 2016; Zacher & Rosing, 2015; Zacher & Wilden, 2015). Hence, the scope of this innovation construct is focused on ideas and improvements surrounding practices at the workplace (and less focused on radical innovations like the development of the television; see also Axtell et al., 2000). To keep our replication as close as possible (cf., Brandt et al., 2014), we designed a task that allowed us to tap into similar aspects of the innovation construct (e.g., Zacher et al., 2016).

In the experimental task, participants have to improve marketing material for an organisation that is trying to promote a 20 year longitudinal study (Appendix C). To carry out

this task, participants receive an existing marketing document that the organisation intends to use for recruitment purposes. This marketing flyer is relatively poorly designed (see Appendix D, i.e., the flyer contains typographical mistakes, words are missing, no formatting, i.e., very small font size etc.); hence, there are various opportunities for improvements (i.e., "finding better ways to do things"). Participants receive relatively narrow task instructions which are to (1) add pictures to the existing marketing flyer, and (2) highlight specific statements within the flyer using different colours. Some may argue that the task merely taps into measuring creativity because participants can focus on adding creative pictures and colourful design features. However, we intentionally chose the narrow task instructions to be able to capture the implementation aspect of innovation. In other words, participants can implement (novel and useful) changes into their work (e.g., reformatting the flyer) that were not part of the work instructions (i.e., "they can find better ways to do things"). Accordingly, we argue that this task reflects central aspects of innovation as defined by Hughes et al. (2018). That is, participants need to identify problems (and opportunities) that are associated with poorly (or well) designed marketing material, they can introduce, adopt or modify new ideas germane to organizational needs (i.e., recruiting and retaining participants) and they practically implement these ideas straightaway in a revised document. Furthermore, we conducted a small pilot study (without leader manipulations) using a sample of full-time working Mturk participants ($N = 19$) who carried out the experimental task (improving the flyer for marketing purposes). In this pretest, we noticed that that participants do indeed implement changes that go beyond the actual task instructions (see Appendix I).

In addition, we checked the validity of the experimental task by correlating a self-report measure of innovative work behaviors (using scales from previous research, i.e., Zacher et al.,

2016, Appendix H) with an external rating of the innovation outcome (i.e., two independent raters rated the uploaded flyers with the innovation outcome scale presented in Appendix G[6]). The objective innovation ratings significantly correlated with participants' self-reported innovation behaviors during the task ($r = .60$, $p = .007$). This result tentatively supports that the experimental task taps into the construct of innovation and will allows us to capture innovative behaviors.

Participants will upload the implementation of their ideas after 30 minutes (to keep this time window constant for all participants, the experiment is programed so that the revised document can only be submitted between 25 min and 30 min.)

Procedure. Participants have access to an external link that directs them to the study, which is set up using the Qualtrics survey platform. Upon clicking on the link, participants are informed about their right to opt out of the study at any time. We guarantee their anonymity and ask participants to give their informed consent before starting the study. In a first step, participants provide information on their demographics (i.e., age, gender, country of residence, English-language proficiency, work hours per week, job role, and industry) as well as on their personality and trait positive affect.

We will check the inclusion criteria and invite those participants for the experiment within a week after taking the first survey. Specifically, those participants will receive a link to the randomized online between-subject experiment. In the main experimental study, all participants first receive the general task instructions (see Appendix C), access to the materials (Appendix D), and are then randomly assigned to a one of the four conditions: (1) leader opening, (2) leader closing, (3) leader ambidexterity, and (4) transformational leadership (see

[6] See Appendix I for examples of flyers with a high and a low innovation rating

Appendix E). In each of the four conditions, participants receive two emails sent from their supervisor: the first time immediately before the task and the second time during the task (i.e., after 15 minutes). We based our decision to divide the task in two periods of 15 minutes whereby the first 15 minutes should be attributed to exploring (via opening behaviors) and the last 15 minutes to exploiting (via closing behaviors) on Farr et al.'s (2003) phase model of innovation. This phase model assumes that exploratory processes are relevant in early phases of an innovation (problem identification and idea generation), whereas late phases require exploitative processes (i.e., idea evaluation and implementation).

The appropriate time periods that are necessary to elicit the proposed positive consequences of opening and closing behaviors are currently unclear from theory. However, three starting points offer some evidence for the fine-grained variations of opening and closing behaviors over time. First, providing support for their dynamic fluctuation on the day level, research has suggested that leadership ambidexterity can vary and exert influence on employees on a daily level (Zacher et al., 2014). Second, the originators of the theory discussed that employee ambidexterity may co-vary with innovation "across even shorter time scales, for example, on an hourly basis" (Rosing et al., 2017, p. 706). Finally, research from economics (e.g., Ederer & Manso, 2009) has provided evidence that short experiments can be suitable to create variations in exploration behavior and innovation. For these reasons, we assume that a timeframe of 30 min is suitable for an experimental test of the theory.

The content of the email contains the manipulations of the different leadership behaviors (Appendix E). Such vignette-based experiments have proven to be as effective as laboratory experiments at evoking responses and can be used to zoom into a "snapshot" of a daily situation

such as leader-follower collaboration on innovation tasks (Aguinis & Bradley, 2014; for a similar approach, see Farh & Chen, 2014).

Following the first email, participants start working on the task. After half of the allocated time, they receive the second email from their supervisor. When the time is over, participants are asked to upload the revised file.

We will incorporate multiple checks to ensure that participants carefully engage in this task. First, in all conditions, participants are informed that their work will be reviewed by the supervisor, who must approve their progress for them to receive payment. Second, participants are only retained in the sample if they respond (in a meaningful way) to their supervisors' emails using an open-ended empty text field ("email to supervisor"). Third, we will include multiple attention and comprehension checks (i.e., instructional manipulation checks, comprehension questions about the experimental materials, cf., Thomas & Clifford, 2017) at the end of the survey.

After uploading the work-output file (which will be rated for innovation), participants report on their exploration and exploitation behaviors during the task. At the end of the survey, participants receive a code to collect their earnings via TurkPrime.

Experimental conditions

Appendix E provides the full material that we will use for each leadership condition.

Leader opening condition. Leader opening has been defined as behaviors that increase variance in followers. Scholars have criticized that constructs should not be defined by their outcomes and that construct definitions should not contain (often incomplete) examples of what is included in a construct (MacKenzie, 2003). Nevertheless, in the absence of a more accurate definition of ambidextrous leadership, we had to rely on items from published scales to experimentally manipulate opening and closing behaviors as expressed in different supervisor

emails (i.e., conditions). Specifically, we created these emails based on the list of opening behaviors provided by Rosing et al. (2011; also used as survey items by Zacher et al., 2016; Zacher & Rosing, 2015; Zacher & Wilden, 2014). Specifically, the supervisor encourages experimentation with different ideas ("I want to strongly encourage you to play with different ideas"), motivates taking risks ("So I encourage you to take a risk"), and allows different ways of accomplishing a task ("think about different ways and methods to make the most of this").

Leader closing condition. Leader closing as expressed in the supervisor email decreases variance in the participant based on the list of closing behaviors provided by Rosing et al. (2011; also used as survey items by Zacher et al., 2016; Zacher & Wilden, 2014; Zacher & Rosing, 2015). Specifically, the supervisor stresses the importance of variance reduction ("I am here to provide you with specific instructions for this task") by monitoring goal attainment ("I regularly check the progress of those workers that I have to supervise"), establishing routines ("If you have established a work routine, I would recommend sticking to it now."), and pointing out adherence to rules ("I believe that close adherence to the rules and the formal task requirements is the best way to be successful here.").

Leader ambidexterity condition. Leader ambidexterity as expressed in the supervisor email is a combination of the opening and closing behaviors provided by Rosing et al. (2011). Specifically, in line with the requirements of the innovation task, the supervisor first increases variance in the participant (i.e., opening) in the first half (e.g., "There is no right or wrong way in doing this…") and then shifts towards variance reduction (i.e., closing) in the second half (e.g. "If you have established a work routine, I would recommend sticking to it now.").

Transformational leadership condition (control). To allow for a fair comparison of our experimental conditions (Cooper & Richardson, 1986), we use transformational leadership as a

control condition. We use transformational leadership for two reasons. First, ambidextrous leadership theory was explicitly introduced as an innovation-specific leadership style that should more accurately predict employee innovation than the heterogeneous findings derived from studies on transformational leadership and innovation (Rosing et al., 2011). Second, two of the studies that we seek to replicate have also used transformational leadership as a control variable (Zacher & Rosing, 2015; Zacher et al., 2016).

To create supervisor emails that express high levels of transformational leadership, we first carefully reviewed published research using transformational leadership vignettes (e.g., Christie, Barlin, & Turner, 2011; Felfe & Schyns, 2006; Hentschel, Braun, Peus, & Frey, 2018). Furthermore, we created the transformational emails by adapting existing transformational leadership vignettes (Christie, Barlin, & Turner, 2011) for our context and by rewording items from published transformational leadership scales (e.g., Rafferty & Griffin, 2004). Specifically, the leader expresses a strong *vision* ("My goal is that our research should create significantly more insights compared to studies from other institutions"), motivates by *inspirational communication* ("This project makes me very proud. I sincerely hope that I can inspire you to feel proud about it as well"), *intellectual stimulation* ("I am encouraging you to think about these problems in new ways"), expresses *support* ("My goal as a leader is to encourage your personal development and to pay attention to your individual needs"), and *personal recognition* ("I will acknowledge your accomplishments when I see outstanding work").

Validity of leadership manipulations (pilot study)

To ensure the validity of the experimental manipulation, we conducted a pilot study containing only the email vignettes and the manipulation-check items. We separated this study from the main experiment in acknowledgement of the recent discussion about the manipulation

check in itself being an intervention that potentially influences participants' subsequent behaviors in the experimental task (e.g., Bless & Burger, 2016; Fayant, Sigall, Lemonnier, Retsin, & Alexopoulos, 2017). We recruited participants for financial compensation (USD 3 for about 15 minutes) using MTurk workers who had an approval rate of more than 90% (Keith et al., 2017) and were located in English-speaking countries (mostly the US). Upon providing their informed consent, participants were randomly assigned to one of the four conditions.

Participants read the emails from the supervisor emails and could use emails to respond to the supervisor (e.g., "You can write an email to shortly discuss the task with your supervisor J.P."). We included this interactive element to make the task more immersive and also as an implicit measure to capture participants' task engagement. We also employed four additional attention check measures and excluded participants who did not answer these checks correctly to ensure that all participants in our final sample had paid careful attention to (and comprehended) the leadership manipulations (see Appendix F for details). We were strict about excluding participants who failed those manipulation checks, because a lack of attention to the supervisor emails would not allow us to detect differences in leadership perceptions (i.e., whether the leader behavior was indeed perceived to be closing, opening, ambidextrous, or transformational). Participants ($N = 75$) had a mean age of 36 years ($SD = 10.2$) and were working on average 37 hours/week; 52% were female.

After reading the supervisor emails, participants rated the extent of opening behavior and closing behavior ("To which extent did the supervisor (J.P) from this task mention or show the following aspects in his/her email?"). We used seven items for each scale (items from Zacher et al., 2016; see also Zacher & Rosing, 2015; Zacher & Wilden, 2014). Opening was measured with the following items, e.g., "allows different ways of accomplishing a task," "encourages

experimentation with different ideas," "motivates to take risks," α =.95). Closing was measured with the following items, e.g., "monitors and controls goal attainment," "takes corrective action," "controls adherence to rules," α = .91). Following the approach by Zacher et al. (2016), we calculated the multiplicative interaction term between the opening scale and the closing scale to obtain a measure of leader ambidexterity.

Participants also rated the extent to which the leader displayed transformational leadership style (using items adapted from Rafferty and Griffin, 2004, α =.95). These items assessed core aspects of transformational leadership like *vision* (e.g., "J.P. mentioned that he/she had a clear understanding of where we are going"), *inspirational communication* (e.g., "…wants to make me proud to be a part of this project"), *intellectual stimulation* (e.g. "… wants to challenge me to rethink some of my basic assumptions about this task."), supportive leadership ("… wants to consider my personal feelings."), and *personal recognition* ("he/she would commend me when I do a better than average job."). To separate the effects of transformational leadership in followers (e.g., "this leader inspires me") from actual behavioral elements of leadership, we added a stem to each item that highlighted the behavioral component of the leadership style "J.P. mentioned in the email that he/she wants to" followed by the items. All items were answered using a 5-point Likert response format (1 = not at all, 2 = a little, 3 = a moderate amount, 4 = a lot, 5 = a great deal; α = .91).

We conducted analyses of variance (ANOVAs) to test if the four conditions resulted in significantly different perceptions of leadership styles. ANOVAs showed that perceptions of opening behaviors differed significantly between the leadership conditions ($F(3, 71) = 31.91$, $p < .001$), perceptions of closing behavior differed significantly between the leadership conditions ($F(3, 71) = 19.28$, $p < .001$), measures of ambidexterity differed significantly between the

conditions ($F(3, 71) = 2.69$, $p = .052$), and perceptions of transformational leadership behaviors also differed significantly between the different conditions ($F(3, 71) = 7.45$, $p < .001$).

Furthermore, we tested specific contrasts to evaluate whether conditions were sufficiently distinct from each other. Contrasts showed that perceptions of opening behaviors were highest in the opening condition ($M = 4.48$, $SD = 0.52$), that is, significantly higher than in the closing condition ($M = 2.06$, $SD = 1.09$, $t(24) = 8.5$, $p < .001$), higher than in the transformational condition ($M = 2.87$, $SD = 1.01$, $t(36) = 4.31$; $p < .001$), and also higher than the ambidexterity condition ($M = 4.06$, SD = 0.68, $t(39) = 2.16$, $p = .037$).

Perceptions of closing behaviors were highest in the closing condition ($M = 3.63$, $SD = 0.68$), that is, significantly higher than in the opening condition ($M = 1.66$, SD = 0.73, $t(35) = 8.40$; $p < .001$), significantly higher than in the transformational condition ($M = 1.95$, $SD = 0.84$, $t(32) = 6.39$, $p < .001$) and also significantly higher than in the ambidexterity condition ($M = 2.36$, SD = 1.02, $t(38) = 4.5$, $p < .001$).

Perceptions of transformational leadership were highest in the transformational condition ($M = 3.49$, $SD = 0.72$), that is, significantly higher than the closing condition ($M = 2.12$, $SD = 1.08$, $t(32) = 4.29$; $p < .001$), significantly higher than in the opening condition ($M = 2.72$, $SD = 0.79$; $t(33) = 2.98$; $p = .005$), and also significantly higher than in the ambidexterity condition ($M = 2.32$, $SD = 0.98$; $t(36) = 4$, $p < .001$).

Finally, the ambidexterity score was highest in the ambidexterity leadership condition ($M = 9.61$, $SD = 4.92$), that is, significantly higher than in the transformational leadership condition ($M = 5.9$, $SD = 3.6$, $t(36) = 2.56$; $p = .015$), and marginally higher than in the opening condition ($M = 7.37$, $SD = 3.24$; $t(39) = 1.70$; $p = .098$) and the closing condition ($M = 7.41$, $SD = 4.20$, $t(38) = 1.5$; $p = .141$).

These results not only show that the different leadership conditions were able to effectively manipulate the different leadership styles, but they also show that the conditions differentiate in nuanced ways between leadership style that share conceptual overlap (e.g., ambidexterity and opening conditions shared 50% of content overlap; opening behaviors and transformational leadership share conceptual elements such as intellectual stimulation and opening; opening and transformational have also shown to correlate in past research, e.g., $r = .67$ in Zacher et al., 2016; $r = .49$ in Zacher & Rosing, 2015).

Finally, we assessed whether the scenarios were perceived as realistic (using two questions from Farh & Chen, 2014: "It is realistic that I might experience a supervisor like J. P.," and "At some point during my career, I will probably encounter a situation like the one described above,") using a 5-point Likert scale (1 = completely disagree, 5 = completely agree, $\alpha = .88$). Participants generally agreed the scenario was realistic ($M = 3.66$, 72% scored ≥ 3) and this perception was not affected by different conditions ($F(3,71) = .40$, $p = .751$). The email responses from participants also indicated that that they were immersed in the scenario (average length per email in characters: $M = 431$; 97% of all participants wrote > 77 characters). In the general feedback section of the survey, participants also commented about J.P. (e.g., "I feel like this is the type of supervisor I would prefer to work with. They give you enough room to hang yourself but sound like they would encourage you to learn from whatever mistake you make." or "The contents were realistic. JP was very focused on the results and consistency."); participants also commented on the interactive elements ("it was an interactive survey where you had to pay attention even to the little details you thought that didn't matter. it was interesting."). These results further support that the experimental conditions were immersive and increase the external validity of this study.

Measures (main study)

Follower exploration, exploitation, and ambidexterity. Participants will rate the extent to which they engaged in exploration and exploitation with 14 items adapted from Zacher et al. (2016) and Mom et al. (2009) using a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). Since these original scales have been used within an organizational field context, we had to reword the items to better align the constructs with the context of the experimental task (see Table 2). Doing so, we carefully paid attention to capture core conceptual features from the definition of individual (non-managerial) ambidexterity. In this definition, exploration encompasses "behaviors related to experimentation, searching for alternative ways to accomplish task, and learning from errors, […] deviat[ing] from routines, trying out something new, and […] not rely[ing] on established knowledge" (Rosing & Zacher, 2017: 696). Hence, exploration is measured with items like "During the task, I focused on strong renewal of the flyer." (Original item: "Focusing on strong renewal of products/services or processes) or "I searched for novel ways to make the flyer more interesting." (Original item: "Searching for new possibilities with respect to my work").

Exploitation is defined as "relying on previous experience, putting things into action, and incrementally improving well-learned actions (…), doing things as they have always been done and relying on existing rules and routines". (Rosing & Zacher, 2017: 696). Hence, examples for exploitation items are "During the task, I focused mainly on carrying out those task activities that were provided in the task description (i.e., adding pictures and colors)." (Original item: "Activities which I clearly know how to conduct") or "Focused on getting the task done as quickly as possible." (Original item: "Activities primarily focused on achieving short-term goals"). The item wording is provided in Table 2. Following Mom et al. (2009) and Zacher et al.

(2016), ambidexterity is operationalized by calculating the multiplicative term of the exploration and exploitation scores.

++++++ Insert Table 2 about here++++++

Innovation outcome. The innovation of the revised marketing flyer will be assessed by two subject matter experts (i.e., researchers / marketing experts who are working on the marketing of the actual 20-year longitudinal study). The raters will be blind to the experimental conditions and independently rate the outcomes submitted by participants using four items (see Appendix G) from Rosing et al. (2018) on a 5-point scale which focuses on the product facet of innovation (i.e., the "innovative outputs implemented") rather than the process facet of innovation (i.e., "behaviors, actions, and cognitive processes that a person (…) engages in when attempting to generate and implement creative ideas", Hughes et al., 2018: 13). Furthermore, this rating scale captures both the creative dimension ("this task outcome is completely novel and does not at all rely on conventional solutions"; "this task outcome is very creative.") and the implementation dimension of innovation ("this task outcome can readily be applied in the 'real world'"; "this task outcome exceeds the quality standards")[7]. After establishing interrater reliability, we will use the average between the two raters as a measure of innovation.

Instrumental variables. As outlined in more detail below, this study makes use of instruments to reduce endogeneity concerns (for the endogenous variables). Particularly, we intend to use the manipulated conditions as instruments (i.e., the opening condition will be an instrument for exploration, while the closing condition will be an instrument for exploitation).

[7] We acknowledge that another conceptual feature of workplace innovation is the promotion/selling of ideas (cf., Hughes et al., 2018). However, this dimension has not been assessed by previous studies on leader ambidexterity theory (e.g., Zacher et al., 2016; Zacher & Rosing, 2015; Zacher & Wilden, 2014). To keep the replication closer to original studies (Brandt et al., 2014), we decided to not include it in our innovation measure.

In addition to this, we will also measure participants' conscientiousness (using 10 items) and openness to experience (using 10 items) from the HEXACO inventory (Ashton & Lee, 2009) and positive trait affect and negative trait affect with five items each from Mackinnon et al. (1999); see Appendix H for the full measures). Although we acknowledge that based on theory (see Figure 1) the relevance condition, in the form of the F-test of the first stage regression, will be less likely to be met with the personality/affect variables than with the manipulated conditions, we nevertheless decided to include them to have some alternative instruments and to reflect previous work. That is, we chose these alternative instruments because follower personality measures (i.e., openness to experience, conscientiousness) and positive affect are exogenous and have been theoretically (Rosing et al., 2011) as well as empirically (Zacher et al., 2016) associated with follower ambidexterity.

We will separate the assessment of the personality and trait affect from the main experiment by using a prescreening procedure (for details see Keith et al., 2017), that is, participants will respond to these instruments (and also demographic information) a week before the actual study and then be re-invited for the actual task.

Analytical strategy

A full overview of the analytical steps that we will perform (including a commented syntax using SPSS and MPLUS as well as key decisions for the results) is provided in Appendix J. We will shortly discuss the most important steps: Only participants who finalize the complete survey, correctly reply to comprehension items (Thomas & Clifford, 2017; Ward & Meade, 2018, for details see Appendix F) are included in the final analysis.

First, we will conduct confirmatory factor analyses to check the robustness of the measurement models underlying our measures of follower exploration and exploitation. We use

Mplus (Muthén & Muthén, 2010) to run a confirmatory factor analysis with two factors (exploration and exploitation), fixing the first loading of each factor at equal to 1.0 (marker variable) and evaluating the goodness of fit of the tested models with model $\chi^2$and its associated p-value. If the χ^2 rejects the model, we will check modification indices for sources of potential misfit and discuss the sources of potential misfit within the discussion section of our study.

Second, we perform an interrater reliability analysis using the innovation outcome ratings to calculate the ICC. The ICC is a statistical index commonly used to estimate reliability because it adjusts for chance agreement and systematic differences between raters (Fleiss & Shrout, 1978); it is therefore a more conservative estimate than the Pearson product-moment correlation. Following Cicchetti (1994), we classify ICCs by cut-off criteria: below .40 = poor, .40–.59 = fair, .60–.74 = good, and .75–1.00 = excellent. In case agreement is ICC ≤ .60, raters will discuss misalignments in which ratings differ and come to an agreement in consultation with the first and/or second authors.

Third, we test the research hypotheses (H1–H4). To test H1–H3, we run ANOVAs and contrasts (using leadership as an independent variable with four conditions: opening, closing, ambidextrous leadership, and transformational leadership) for each of the three dependent variables (follower exploration, follower exploitation, and innovation). A detailed description of the predicted results is further provided in Table 3. To illustrate, the first row shows that for Hypothesis 1, we expect a significant F-value for the one-factorial ANOVA (with four leadership conditions) when using exploration as a dependent variable. Furthermore, we expect that contrasts between conditions to show significant differences between the opening and closing condition (with higher exploration values in the opening condition).

++++++ Insert Table 3 about here++++++

To test H4, we will run multiple regression analysis using exploration, exploitation, and their multiplicative interaction (follower ambidexterity) as predictors of innovation. Because follower ambidexterity is assessed via self-reports and not experimentally manipulated, it is possible that the error terms across equations are still correlated due to the omission of common causes of both variables (Shaver, 2005). This could result in an inconsistent estimate of the relationship between the follower measures (i.e., exploration, exploitation) and innovation (Antonakis, et al., 2010). To avoid this problem, we apply a two-stage least squares (2SLS) regression. 2SLS allows to obtain a consistent estimate of the coefficient (i.e., relationship between follower ambidexterity and the outcome variable) by using an instrumental variables approach. Instrumental variables must be exogenous (i.e., they should vary randomly in nature or be experimentally manipulated), should satisfy the exclusion condition (i.e., they should not be direct predictors of the outcome variable, innovation, beyond their effect on the endogenous variable, that is, follower exploration and exploitation behaviors), and they should be strong predictors of follower ambidexterity (i.e., the relevance condition; cf., Antonakis et al., 2010; Ketokivi & McIntosh, 2017)."

We intend to use the manipulated conditions as instruments. The relationship between the instrumental variables and ambidexterity (i.e., exploration, X_1, and exploitation, X_2) is first estimated, which provides predicted values of X_1 and X_2. We will test the appropriateness of these instruments with the F-test of this regression (i.e., > 10, Staiger & Stock, 1997). If the manipulated conditions pass the test, we regress our dependent variables on the predicted values of ambidextrous follower behaviors based on the estimates of the first stage. In case the manipulated conditions turn out to be insufficient predictors of ambidextrous behaviors, we drop

them from the analysis and test whether other instruments (i.e., personality, trait positive affect) pass the test.

Study 2: Laboratory Experiment

The purpose of Study 2 is to replicate the findings from Study 1, and to increase the external validity by using a more complex representation of behavior than that allowed in a vignette study. That is, our second study tests the leader ambidexterity model in an experimental context that more realistically mimics the fine-grained dynamics of leadership ambidexterity theory. To do so, we will manipulate leadership behaviors by using a trained actor who will instruct participants to conduct the same experimental task as in Study 1. To ensure a high standardization of the manipulation, we videotape the actors' behaviors (Antonakis, d'Adda, Weber, & Zehnder, 2015). Manipulating leadership style through the behaviors of an actor provides further opportunity to depict more realistically the nature of the leadership styles (Avolio, Reichard, Hannah, & Walumbwa, Chan, 2009; Podsakoff, Podsakoff, MacKenzie, & Klinger, 2013). Furthermore, the study allows for a more subtle manipulation of different leadership styles and, thus, a stricter test of the model.

Methods

The study is subject to approval by the institutional review board of the XX University prior to conducting the study. Study data will be published in anonymized form on the Open Science Framework (openscience.org).

Sample. The sample size calculations for Study 1 were conducted using effect sizes from non-experimental field studies (see Table 1). There are strong arguments that a latent (i.e., true) moderator effect is often severely underestimated by correlational field studies and that experimental designs can reveal substantially larger moderation effects (for details about these

arguments see Murphy & Russel, 2017). Based on these arguments, we will re-run calculations for the required sample size for Study 2 based on the effect sizes obtained from Study 1. The results of this analysis will determine the number of recruited participants. In case this number is larger than 250, we conduct the experiments in two labs (i.e., one in XX and one in the XX[8]) in order to manage data collection (we will control for location).

*Design***.** We will use an experimental design with four leader conditions (opening, closing, ambidextrous leadership, and transformational leadership). The participants will work on the same task as described in Study 1. In contrast to Study 1, participants will receive instructions via two video-messages. The use of video-messages allows us to deliberately manipulate the independent variable, that is, the behavior of the leader in a standardized way but also to create a more immersive (and natural) scenario (Antonakis et al., 2015; Podsakoff et al., 2013). In other words, video-based manipulations constitute very effective ways to assess construct the theoretical validity of a construct (Podsakoff et al., 2013) while maintaining high levels of both internal and external validity. Video-based manipulation of leader behaviors have high external validity because the method engages participants' senses more fully, is more lifelike, and provide a greater amount of "natural noise" in each scenario (Aguinis & Bradley, 2014). At the same time, this method has high internal validity because the focal independent variable (i.e., leadership behavior) is kept constant and controlled by the researcher in each video stimuli.

A professional actor will be hired to enact the leadership conditions [(a) opening behavior, (b) closing behavior, (c) ambidextrous leadership, (d) transformational leadership]. To do this, the actor will be given the email transcripts (used in Study 1, see Appendix E) to portray

[8] Anonymized for review purposes

the leader in two video messages per condition. To minimize effects of the actor (i.e. physical appearance, attractiveness, gender), the same actor will be used across all four conditions (cf., Podsakoff et al., 2013). The actor will be instructed to portray only differences in those leadership styles while maintaining contaminating factors (e.g., non-verbal cues, body movements and orientation) constant. Furthermore, the visual perspective and film editing will be kept constant across conditions and we will pay attention that the actor uses exactly the same wording of the emails (see Appendix E that provides the transcript for his role).

Procedure. Before arriving, participants have to provide their informed consent for taking part in the study and fill in a short survey on their demographic information, personality, and trait positive and negative affect.

Upon arrival, participants will be instructed that they will be working out marketing material for a real research project (i.e., longitudinal study). Following this, participants are randomly allocated to one of the four leadership conditions. The paradigm was modeled after lab-based experimental leadership research methods using video-based messages (e.g., Damen, van Knippeberg, & van Knippenberg, 2008; Venus, Stam, & Van Knippenberg, 2010). Participants will be placed inside cubicles behind a computer and informed that they will receive two videotaped messages by their supervisor. In the video-message, the supervisor provides specific instructions for the task using a different leadership style in each of the conditions. After 30 minutes, participants have to submit the final output. After submission, participants will indicate their exploration and exploitation activities during the task. At the end, participants will be fully debriefed and compensated.

Experimental manipulation

Opening, closing, ambidexterity, and transformational leadership conditions. The four different conditions in the video messages will all be based on the same email transcripts that we used for Study 1 (see Appendix E).

Validity checks for leadership conditions. While the content validity of the written leadership vignettes has already been assessed as part of the pre-tests in Study 1, we will assess the content validity of the video-recorded messages using the same procedure as described in the pilot of Study 1. That is, after having produced the video-messages, we will present the materials to a separate sample (N = 75) and test whether these conditions are rated significantly different on validated scales of leader opening, leader closing, and transformational leadership (same scales as used for Study 1). Based on these analyses, we might have to re-fine the recorded video manipulations by advising the actor to accentuate different leadership behaviors differently until the stimulus materials obtains sufficiently high levels of construct validity (i.e., each leadership condition needs to have a mean rating of $M \geq 3.5$ for the respective leadership style on a 5 point Likert extent-scale with 1 = not at all, 2 = a little, 3 = moderate amount, 4 = a lot, 5 = a great deal) and discriminant validity (i.e., the ratings of the focal leadership style needs to be at least 0.5 points higher in the corresponding leadership condition in comparison to the other leadership conditions; e.g., the rating of opening behavior needs to be 0.5 points higher in the opening condition than in the transformational condition; this difference needs to be significant).

Control (leadership behaviors). To rule out that the video-messages of the leadership conditions do not differently affect the work process of participants in each condition, we will also measure perceived interruptions in each condition. We will use an adapted scale from Sonnentag, Reinecke, Mata and Vorderer (2017), that is, participants will indicate their

agreement to the following items using a 5-point Likert scale (1 = fully disagree; 5 = fully agree): “Incoming video-message kept me from doing my job,” “The video-messages have reached me at inconvenient moments,” and “The video-messages disturbed me in doing my work.”

++++++ Insert Table 3 about here++++++

Measures

Follower exploration, exploitation, and ambidexterity. Exploration behavior and exploitation behavior will be captured using the same measures as in Study 1.

Instrumental variables. Similar to Study 1, we intend to use the manipulated conditions as instruments. Furthermore, to have some alternative potential instruments, we will measure participants’ conscientiousness and openness to experience from the HEXACO inventory (Asthon & Lee, 2009), as well as their trait positive and negative affect (Mackinnon et al., 1999).

Attention/comprehension checks. We will only include participants in our study that will pass comprehension checks at the end of the study, we will use the same measures of attention and comprehension test (for examples, see Appendix F).

Manipulation checks (leadership behaviors). Finally, to verify our manipulations of the different leader behaviors, participants will also rate the leader from the video (at the end of the task and after having provided ratings on all other measures). We will be using the same items for opening leadership, closing leadership, and transformational leadership as in Study 1 to evaluate whether participants perceived the confederates’ leadership behaviors as intended.

Innovation outcomes. Similarly as in Study 1, subject experts blind to the conditions will rate the innovativeness of the solution developed by the participants using four items (Rosing et al., 2018; see also Appendix G).

Analytical Strategy

We will use the same analytical approach as outlined in Study 1. A full overview of the analytical steps is also provided in Appendix J

Discussion

This conceptual replication study intends to replicate the findings on the effects of ambidextrous leadership on followers' ambidextrous behaviors and innovation through an experimental study conducted in two different settings, namely a crowdsourcing online sample (similar to the recruitment process of one of the largest samples in previous leadership ambidexterity research, e.g., Zacher et al., 2016) and in the lab (not used previously). The replication of each hypothesis will be considered successful if we find significant results (as outlined in table 3). In case we do not find any significant relationships, we will carefully discuss potential reasons (e.g., experimental design, analytical approach, and/or different operationalization). In any case, the study is important in contributing towards a more rigorous leadership research, and in helping to advance a better theoretical understanding of leadership in the context of innovation.

References

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, *17*, 351–371.

Aguinis, H., & Gottfredson, R. K. (2010). Best-practice recommendations for estimating interaction effects using moderated multiple regression. *Journal of Organizational Behavior, 31*, 776 –786.

Aguinis, H., Gottfredson, R. K., & Wright, T. A. (2011). Best-practice recommendations for estimating interaction effects using meta-analysis. *Journal of Organizational Behavior*, *32*, 1033–1043.

Alghamdi, F. (2018). Ambidextrous leadership, ambidextrous employee, and the interaction between ambidextrous leadership and employee innovative performance. *Journal of Innovation and Entrepreneurship*, *7*(1), 1–14.

Antonakis, J., Bastardoz, N., Jacquart, P., & Shamir, B. (2016). Charisma: An ill-defined and ill-measured gift. *Annual Review of Organizational Psychology and Organizational Behavior*, *3*, 293–319.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly, 21,* 1086–1120.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2014). Causality and endogeneity: Problems and solutions. In D.V. Day (Ed.), *The Oxford Handbook of Leadership and Organizations* (pp. 93–117). New York: Oxford University Press.

Artz, K. W., Norman, P. M., Hatfield, D. E., & Cardinal, L. B. (2010). A longitudinal study of the impact of R&D, patents, and product innovation on firm performance. *Journal of Product Innovation Management*, *27*, 725–740.

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340–345.

Avolio, B. J., Reichard, R. J., Hannah, S. T., Walumbwa, F. O., & Chan, A. (2009). A meta-analytic review of leadership impact research: Experimental and quasi-experimental studies. *The Leadership Quarterly*, *20*, 764–784.

Axtell, C. M., Holman, D. J., Unsworth, K. L., Wall, T. D., Waterson, P. E., & Harrington, E. (2000). Shopfloor innovation: Facilitating the suggestion and implementation of ideas. *Journal of Occupational and Organizational Psychology*, *73*, 265 –285.

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior. *Perspectives on Psychological Science*, *2*, 396–403.

Behrendt, P., Matz, S., & Göritz, A. S. (2017). An integrative model of leadership behavior. *The Leadership Quarterly, 28*, 229–244.

Bledow, R., Frese, M., Anderson, N., Erez, M., & Farr, J. (2009). A dialectic perspective on innovation: Conflicting demands, multiple pathways, and ambidexterity. *Industrial and Organizational Psychology*, *2,* 305–337.

Bless, H., & Burger, A. M. (2016). A closer look at social psychologists' silver bullet: Inevitable and evitable side effects of the experimental approach. *Perspectives on Psychological Science, 11,* 296–308.

Camisón-Zornoza, C., Lapiedra-Alcamí, R., Segarra-Ciprés, M., & Boronat-Navarro, M. (2004).

A meta-analysis of innovation and organizational size. *Organization Studies*, *25*, 331–361.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior, 29,* 2156–2160.

Cheung, J. H., Burns, D. K., Sinclair, R. R., & Sliter, M. (2017). Amazon Mechanical Turk in organizational psychology: An evaluation and practical recommendations. *Journal of Business and Psychology*, *32*(4), 347–361.

Christie, A., Barling, J., & Turner, N. (2011). Pseudo-transformational leadership: Model specification and outcomes. *Journal of Applied Social Psychology*, *41*(12), 2943–2984.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290.

Cohen, J. E. (1988). *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cooper, W. H., & Richardson, A. J. (1986). Unfair comparisons. *Journal of Applied Psychology, 71*(2), 179–184.

Damen, F., Van Knippenberg, B., & Van Knippenberg, D. (2008). Affective match in leadership: Leader emotional displays, follower positive affect, and follower performance. *Journal of Applied Social Psychology, 38*, 868–902.

David Eccles School of Business (2018). Certificate in Innovative Leadership. Retrieved from http://eccles.utah.edu/programs/executive-education/certificates/#innovative-leadership on 1st of June 2018

Day, D. V., & Antonakis, J. (2012). Leadership: Past, present, and future. In D. V. Day, & J.

Antonakis (Eds.), *The nature of leadership* (pp. 3–25) (2nd ed.). Los Angeles, CA: Sage

Denti, L., & Hemlin, S. (2012). Leadership and innovation in organizations: A systematic review of factors that mediate or moderate the relationship. *International Journal of Innovation Management*, *16,* 1240007-1–1240007-20

Ederer, F., & Manso, G. (2013). Is pay for performance detrimental to innovation?. *Management Science*, *59*(7), 1496-1513.

Farr, J. L., Sin, H.-P., & Tesluk, P. E. (2003). Knowledge management processes and work group innovation. In L. V. Shavinina (Ed.), *The international handbook on innovation* (pp. 574–586). New York, NY: Elsevier Science.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191

Fayant, M. P., Sigall, H., Lemonnier, A., Retsin, E., & Alexopoulos, T. (2017). On the limitations of manipulation checks: An obstacle toward cumulative science. *International Review of Social Psychology*, 30, 125–130.

Farh, C. I., & Chen, Z. (2014). Beyond the individual victim: Multilevel consequences of abusive supervision in teams. *Journal of Applied Psychology*, *99*, 1074–1095.

Felfe, J., & Schyns, B. (2006). Personality and the perception of transformational leadership: The impact of extraversion, neuroticism, personal need for structure, and occupational self-efficacy. *Journal of Applied Social Psychology*, *36*, 708–739.

Fleiss, J. L., & Shrout, P. E. (1978). Approximate interval estimation for a certain intraclass correlation coefficient. *Psychometrika*, *43*, 259–262.

Gibson, C. B., & Birkinshaw, J. (2004). The antecedents, consequences, and mediating role of

organizational ambidexterity. *Academy of Management Journal, 47*, 209–226.

Hentschel, T., Braun, S., Peus, C., & Frey, D. (2018). The communality-bonus effect for male transformational leaders–leadership style, gender, and promotability. *European Journal of Work and Organizational Psychology*, *27*, 112–125.

Hill, A. D., White, M. A., & Wallace, J. C. (2014). Unobtrusive measurement of psychological constructs in organizational research. *Organizational Psychology Review*, *4*, 148–174.

Hughes, D. J., Lee, A., Tian, A. W., Newman, A., & Legood, A. (2018). Leadership, creativity, and innovation: A critical review and practical recommendations. *The Leadership Quarterly, 29,* 549–569.

Hülsheger, U. R., Anderson, N., & Salgado, J. F. (2009). Team-level predictors of innovation at work: A comprehensive meta-analysis spanning three decades of research. *Journal of Applied Psychology, 94*, 1128–1145.

Keith, M. G., Tay, L., & Harms, P. D. (2017). Systems perspective of Amazon Mechanical Turk for organizational research: Review and recommendations. *Frontiers in Psychology*, *8*, 1359.

Keller, T., & Weibler, J. (2015). What it takes and costs to be an ambidextrous manager: Linking leadership and cognitive strain to balancing exploration and exploitation. *Journal of Leadership & Organizational Studies*, *22*, 54–71.

Ketokivi, M., & McIntosh, C. N. (2017). Addressing the endogeneity dilemma in operations management research: Theoretical, empirical, and pragmatic considerations. *Journal of Operations Management*, *52*, 1–14.

Kinni , T. (2016). How to become an ambidextrous leader. Retrieved from *strategy+business Blog*, https://www.strategy-business.com/blog/How-to-Become-an-Ambidextrous-

Leader?gko=91b76 on 1st of June 2018

Kline, R. (2011). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

MacKenzie, S. B. (2003). The dangers of poor construct conceptualization. *Journal of the Academy of Marketing Science, 31*(3), 323-326.

Mackinnon, A., Jorm, A. F., Christensen, H., Korten, A. E., Jacomb, P. A., & Rodgers, B. (1999). A short form of the Positive and Negative Affect Schedule: Evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual differences*, *27*(3), 405–416.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7*, 537–542.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2,* 71–87.

Mom, T. J., Van Den Bosch, F. A., & Volberda, H. W. (2009). Understanding variation in managers' ambidexterity: Investigating direct and interaction effects of formal structural and personal coordination mechanisms. *Organization Science*, *20*, 812–828.

Mumford, M. D., Scott, G. M., Gaddis, B., & Strange, J. M. (2002). Leading creative people: Orchestrating expertise and relationships. *The Leadership Quarterly*, *13,* 705–750.

Murphy, K. R., & Russell, C. J. (2017). Mend it or end it: Redirecting the search for interactions in the organizational sciences. *Organizational Research Methods*, *20*, 549–573.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

O'Boyle, E., Banks, G. C., Carter, K., Walter, S., & Yuan, Z. (2018). A 20-year review of

outcome reporting bias in moderated multiple regression. *Journal of Business and Psychology*, 1–19.

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science, 7,* 657–660.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63,* 539–569.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12,* 531–544.

Podsakoff, N. P., Podsakoff, P. M., MacKenzie, S. B., & Klinger, R. L. (2013). Are we really measuring what we say we're measuring? Using video techniques to supplement traditional construct validation procedures. *Journal of Applied Psychology*, *98*, 99–113.

Rafferty, A. E., & Griffin, M. A. (2004). Dimensions of transformational leadership: Conceptual and empirical extensions. *The Leadership Quarterly*, *15*(3), 329–354.

Rosing, K., Bledow, R., Frese, M., Baytalskaya, N., Johnson Lascano, J., & Farr, J. L. (2018). The temporal pattern of creativity and implementation in teams. *Journal of Occupational and Organizational Psychology*, *91*(4), 798–822. doi:10.1111/joop.12226

Rosing, K., Frese, M., & Bausch, A. (2011). Explaining the heterogeneity of the leadership-innovation relationship: ambidextrous leadership. *The Leadership Quarterly, 22,* 956–974.

Rosing, K., Rosenbusch, N., & Frese, M. (2010). Ambidextrous leadership in the innovation process. In A. Gerybadze, U. Hommel, H. W. Reiners & D. Thomaschewski (Eds.), *Innovation and International Corporate Growth* (pp. 191–204). Berlin: Springer.

Rosing, K., & Zacher, H. (2017). Individual ambidexterity: The duality of exploration and exploitation and its relationship with innovative performance. *European Journal of Work and Organizational Psychology*, *26*(5), 694–709.

Rubera, G., & Kirca, A. H. (2012). Firm innovativeness and its performance outcomes: A meta-analytic review and theoretical integration. *Journal of Marketing, 76,* 130–147.

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f2, a measure of local effect size, from PROC MIXED. *Frontiers in psychology*, *3*, 111.

Shaver, J. M. (2005). Testing for mediating variables in management research: Concerns, implications, and alternative strategies. *Journal of Management*, *31*(3), 330–353.

Simonsohn, U. (2013). Evaluating replication results. Available at SSRN: http://ssrn.com/abstract=2259879

Sonnentag, S., Reinecke, L., Mata, J., & Vorderer, P. (2018). Feeling interrupted - being responsive: How online messages relate to affect at work. *Journal of Organizational Behaviour, 39*(3), 369-383.

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*, 557–586.

Stock, R. M., von Hippel, E., & Gillert, N. L. (2016). Impacts of personality traits on consumer innovation success. *Research Policy*, *45*, 757–769.

Stock, R. M., Zacharias, N. A., & Schnellbaecher, A. (2017). How do strategy and leadership styles jointly affect co-development and its innovation outcomes? *Journal of Product Innovation Management*, *34*, 201–222.

Thomas, K. A., & Clifford, S. (2017). Validity and Mechanical Turk: An assessment of

exclusion methods and interactive experiments. *Computers in Human Behavior*, *77*, 184-197.

Tushman, M. L., & O'Reilly, C. A. (1997). *Winning through innovation*. Cambridge: Harvard Business School Press.

Tushman, M. L., Smith, W. K., & Binns, A. (2011). The ambidextrous CEO. *Harvard Business Review, 89*, 74–80.

Venus, M., Stam, D., & Van Knippenberg, D. (2013). Leader emotion as a catalyst of effective leader communication of visions, value-laden messages, and goals. *Organizational Behavior and Human Decision Processes, 122*, 53-68.

Ward, M. K., & Meade, A. W. (2018). Applying social psychology to prevent careless responding during online surveys. *Applied Psychology*, *67*, 231–263.

West, M. A., & Farr, J. L. (1990). *Innovation and creativity at work: Psychological and organizational strategies.* Chichester: Wiley.

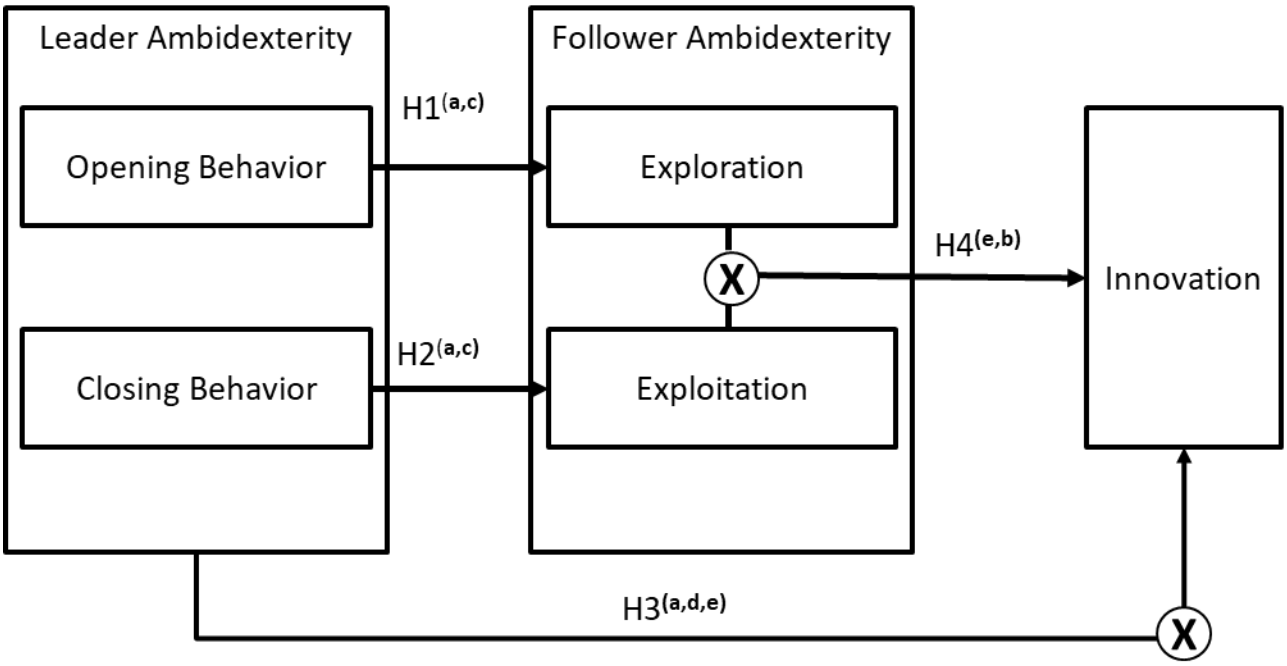
Zaccaro, S. J., & Horn, Z. N. (2003). Leadership theory and practice: Fostering an effective symbiosis. *The Leadership Quarterly*, *14*, 769–806.

Zacher, H., Robinson, A. J., & Rosing, K. (2016). Ambidextrous leadership and employees' self-reported innovative performance: The role of exploration and exploitation behaviors. *The Journal of Creative Behavior*, *50*, 24–46.

Zacher, H., & Rosing, K. (2015). Ambidextrous leadership and team innovation. *Leadership & Organization Development Journal*, *36*, 54–68.

Zacher, H., & Wilden, R. G. (2014). A daily diary study on ambidextrous leadership and self-reported employee innovation. *Journal Of Occupational & Organizational Psychology, 87,* 813–820.

Figure 1. Model of leader ambidexterity for innovation (including empirical support for model paths).



Note: Empirical support for each hypothesis is provided in brackets; superscript letters = direct support
a = Alghamdi (2018)
b = Rosing & Zacher (2017)
c = Zacher et al. (2016)
d = Zacher & Rosing (2015)
e = Zacher & Wilden (2014)

Table 1. Literature review outlining which paths of ambidextrous leadership theory have been tested how in previous research.

Publication	Sample	Design	Control Variables	Measures (IV)	Measures (DV)	Findings	Support of Model Paths
Alghamdi (2018) *Journal of Innovation and Entrepreneurship*	• 147 faculty members working in the Albaha province (Saudi Arabia)	• Cross-sectional • Self-report survey • Single source (from employees) • Correlational	• *Leader:* age, gender, educational level, faculty position	• Employees rated perceived supervisors opening (7 items) and closing behaviors (6 items)[a] • Leader ambidexterity = opening*closing	• Employee-rated exploration (5 items) and exploitation (6 items)[b] • Employee rated innovation (4 items)[c]	• Opening behaviors → employee exploration • Closing behaviors → employee exploitation • Opening*Closing leader behaviors → Employee innovation	• H1, H2, H3
Rosing & Zacher (2017) *European Journal of Work and Organizational Psychology*	• Study 1 (weekly study): 59 employees in Australia • Study 2 (daily study): 37 employees working in creative industries in northern Germany	• 2 diary studies: Study 1 (weekly, six waves); Study 2 (daily, 5 waves) • Single-source (all employee rated)	• *Employee:* Positive weekly affect, Negative weekly affect, innovation requirements, age, gender, education	• Employee-rated exploration (5 items) and exploitation (6 items)[b]	• Employee rated innovative work performance (6 items)	• Employee ambidexterity → employee innovative work performance	• H4

Zacher et al. (2016) *The Journal of Creative Behavior*	• 388 employees • Recruited via MTurk (United States)	• Cross-sectional • Single-source survey (from employees) • Between-subjects • Correlational	• *Leader:* Transf. leadership, transact. Leadership • *Employee*: open., conscient., trait positive affect	• Perceived leader opening (7 items) and closing behaviors (6 items)[a] • Self-rated exploration (5 items) and exploitation (4 items)[b]	• Self-rated innovation performance (4 items)[c] • Self-rated exploration (5 items) and exploitation (4 items)[b]	• Opening behaviors → Employee exploration • Closing behaviors → Employee exploitation • Employee exploration, exploitation, and exploration*exploitation (controlling for opening and closing behaviors)→ self-reported innovation	• H1, H2, H4
Zacher & Rosing (2015) *Leadership & Organization Development Journal*	• 33 team leaders and 90 of their employees from Australia	• Cross-sectional • Dual source (ratings from leaders and employees) • team-level analysis • Between-team comparison • Correlational	• *Leader:* Transf. Leadership • *Team:* General team success	• Employees rated perceived leader opening (7 items) and closing behaviors (7 items) • Controls: Transformational leadership	• Team leaders rated team innovative performance (4 items)[c] • Team leaders rated team success (1 item)	• Opening behaviors → team innovation • Closing behaviors (→X) team innovation • Opening*Closing leader behaviors → team innovation	• H3
Zacher & Wilden (2014) *Journal of Occupational and Orga-*	• 113 employees • Convenience sampling, recruited through personal	• Diary study (within-person and between-person) • Single-source self-report	• *Leader*: Intellectual stimulation, • *Employee:* Positive affect, job autonomy	• Perceived leader opening (4 items) and closing behaviors (4 items) [a]	• Self-rated innovation performance (4 items)[c]	• Opening behaviors → Daily self-reported innovation • Closing behaviors (→X) daily innovation	• H3

nizational Psychology	contacts and research participant pools (no information about nationality) • Baseline survey & daily survey (five workdays)	survey (from employees) • Correlational	• Opening*Closing behaviors → daily innovative performance

Notes. To reduce complexity of this table, we only report study variables in each study that are central to the aims of this replication study. transf. = transformational, transact. = transactional, open. = openness to experience, conscient. = conscientiousness, IV = predictor variables, DV = dependent variables, → means "positively predicted", →*X* means "did not have a significant effect", * = "Multiplicative interaction between two variables", [a] = items based on behavioral descriptions from Rosing et al. (2011), [b] = items used from Mom et al. (2009), [c] = items used from Welbourne (1998)

Table 2. Conceptual definitions of exploration/exploitation, items from field research, and adapted items for the experimental task

Construct definition		Managerial (Mom et al., 2009)	Employee (non-managerial) (Zacher et al., 2016).	Items adapted items for the experiment
Individual exploration: "behaviors related to experimentation, searching for alternative ways to accomplish task, and learning from errors. When exploring, individuals deviate from routines, try out something new, and do not rely on established knowledge." (Zacher & Rosing, 2017: 696)		*To what extent did you, last year, engage in work related activities that can be characterized as follows.*	*Rate the extent to which you engaged in the following activities at work.*	*"During the task,*
	1	Searching for new possibilities with respect to products/services, processes, or markets.	Searching for new possibilities with respect to my work	I searched for novel ways to make the flyer more interesting.
	2	Evaluating diverse options with respect to products/services, processes, or markets	Evaluating diverse options with respect to my work	I evaluated diverse options with respect to the flyer.
	3	Focusing on strong renewal of products/services or processes	Focusing on strong renewal of products/services or processes	I focused on strong renewal of the flyer.
	4	Activities requiring quite some adaptability of you	Activities requiring me to be adaptable	I had to be adaptable.
	5	Activities requiring you to learn new skills or knowledge	Activities requiring me to learn new skills or knowledge	I was trying to learn something new.
	6	Activities that are not (yet) clearly existing company policy	--	I engaged in activities that were not formally required by the task description
	7	Activities of which the associated yields or costs are currently unclear	--	I tried to experiment with different methods to reach the goal.
Individual exploitation: "encompasses relying on previous experience, putting things into action, and incrementally improving well-learned actions. Exploitation involves doing things as they have always been done and relying on existing rules and routines. (Zacher & Rosing, 2017: 696)	1	Activities which serve existing (internal) customers with existing services/products	Activities which serve existing customers with existing products/ services	I maintained the existing format and existing text of the flyer draft version.
	2	Activities of which it is clear to you how to conduct them	Activities which I clearly know how to conduct	Focused mainly on carrying out those task activities that were provided in the task description (i.e., adding pictures and colors)
	3	Activities primarily focused on achieving short term goals	Activities primarily focused on achieving short-term goals	Focused on getting the task done as quickly as possible.
	4	Activities which you can properly conduct by using your present knowledge	Activities I can properly conduct using my existing knowledge	I only conducted those activities which I knew how to conduct
	5	Activities which clearly fit into existing company policy	Activities which clearly fit into existing company policy	I strictly adhered to the rules and fulfillment of task requirements.
	6	Activities which you carry out as if it were routine	--	Focused on implementing those things that were required.
	7	Activities of which a lot of experience has been accumulated by yourself	Activities in which I have accumulated a lot of experience	

Table 3. Analytical approach and predicted differences for the experimental task

Nr.	Hypothesis	Expected ANOVA results (in the experiment)	Expected contrasts in the experiment
1	*Opening leader behaviors positively predict follower explorative behaviors.*	Significant F-value ($p < .05$) with *follower exploration* as dependent variable and leadership style as independent variable (four conditions: opening, closing, ambidextrous, transformational)	*Dependent variable: Follower exploration* Contrasting specific conditions: Opening condition > Closing condition ($p < .05$) Opening condition > Transformational condition ($p < .05$) Opening condition > or = Ambidexterity condition (p = na)[a]
2	*Closing leader behaviors positively predict follower exploitative behaviors.*	Significant F-value ($p < .05$) with *follower exploitation* as dependent variable and leadership style as independent variable (four conditions: opening, closing, ambidextrous, transformational)	*Dependent variable: Follower exploitation* Contrasting specific conditions: Closing condition > Opening condition ($p < .05$) Closing condition > Transformational condition ($p < .05$) Closing condition > or = Ambidexterity condition (p = na)[a]
3	*The interaction of leader opening and closing behaviors (i.e., leader ambidexterity) positively predicts employee innovation, insofar that innovation is highest when both leader opening and leader closing behaviors are high.*	Significant F-value ($p < .05$) with *innovation outcome* as dependent variable and leadership style as independent variable (four conditions: opening, closing, ambidextrous, transformational)	*Dependent variable: External ratings of innovation outcomes* Contrasting specific conditions: Ambidextrous condition > Closing condition ($p < .05$) Ambidextrous condition > Opening condition ($p < .05$) Ambidextrous condition > Transformational condition ($p < .05$)
4	*The interplay of follower exploration and exploitation (i.e., follower ambidexterity) positively predicts innovation outcomes, insofar that innovation is highest when both follower exploration and exploitation behaviors are high.*	Significant F-value ($p < .05$) for the regression model with *innovation outcome* as dependent variable and follower activities as independent variables (three variables: exploration, exploitation, follower ambidexterity)	*Dependent variable (Y): External ratings of innovation outcomes* X1 = Exploration† X2 = Exploitation† X3 = Follower ambidexterity (exploration*exploitation) Multiple regression $Y = B + \beta_1 X1 + \beta_2 X2 + \beta_3 X3$

Note. [a] =Leadership ambidexterity theory (and empirical research) do not allow to make specific predictions about this effect; † = in a first stage, both the endogenous variables X1 and X2 will regressed on the instrumental variables (i.e., three dummy variables for the different leadership conditions) to obtain predicted values of X1 and X2 in order to purge endogeneity biases.

Appendix A

To obtain a measure of the local effect sizes for the main effects (H1, H2) within the theoretical model (Figure 1), we relied on the Pearson correlation *r* which is recommended when they are reported within primary studies (Aguinis, Gottfredson, & Wright, 2011).

To obtain a measure of the local effect sizes for the moderation effects (H3, H4) within the model, we calculated Cohen's f^2 (Cohen, 1988[9]) which is recommended when estimating the local effect size from studies using *moderated* multiple regressions (Aguinis & Gottfredson, 2010; O'Boyle, Banks, Carter, Walter, & Yuan, 2018; Selya, Rose, Hedeker & Mermelstein, 2012). The effect size f^2 can be calculated based on explained variance the by using this formula:

$$f^2 = \frac{(R^2_{AB} - R^2_A)}{(1 - R^2_{AB})} = \frac{\Delta R^2}{(1 - R^2_{AB})}$$

In this formula, *B* is the moderator variable (i.e., x*z), *A* is the set of all other variables in the regression model (i.e., x, z, and control variables). R^2_{AB} is the proportion of variance accounted for by *A* and *B* together (relative to a model with no predictors) and R^2_A is the proportion of variance accounted for by *A* (relative to a model with no predictors). Thus the local effect size f^2 reflects the proportion of variance uniquely accounted for by the moderator effect, over and above all other variables (Aguinis & Gottfredson, 2010; O'Boyle et al, 2018).

Protocol of power analyses using G*Power for

H1: Opening leader behaviors positively predict follower explorative behaviors.

Exact - Correlation: Bivariate normal model

Options:	exact distribution		
Analysis:	A priori: Compute required sample size		
Input:	Tail(s)	=	Two
Correlation ρ H1	= 0.41		
α err prob	= 0.05		
Power (1-β err prob)	= 0.95		
Correlation ρ H0	= 0		
Output:	Lower critical r	=	-0.2335230
Upper critical r	= 0.2335230		
Total sample size	= 71		
Actual power	= 0.9511882		

H2: Closing leader behaviors positively predict follower exploitative behaviors.

Exact - Correlation: Bivariate normal model

Options:	exact distribution

[9] According to Cohen (1988) $f^2 \geq 0.02$, $f^2 \geq 0.15$, and $f^2 \geq 0.35$ represent small, medium, and large effect sizes

Analysis:	A priori: Compute required sample size		
Input:	Tail(s)	=	Two
Correlation ρ H1	= 0.21		
α err prob	= 0.05		
Power (1-β err prob)	= 0.95		
Correlation ρ H0	= 0		
Output:	Lower critical r	=	-0.1154067
Upper critical r	= 0.1154067		
Total sample size	= 289		
Actual power	= 0.9506316		

H3: The interaction of leader opening and closing behaviors (i.e., leader ambidexterity) positively predicts employee innovation, insofar that innovation is highest when both leader opening and leader closing behaviors are high.

F tests - Linear multiple regression: Fixed model, R^2 increase

Analysis:	A priori: Compute required sample size		
Input:	Effect size f^2	=	0.05
α err prob	= 0.05		
Power (1-β err prob)	= 0.95		
Number of tested predictors	= 1		
Total number of predictors	= 4		
Output:	Noncentrality parameter λ	=	13.1000000
Critical F	= 3.8778963		
Numerator df	= 1		
Denominator df	= 257		
Total sample size	= 262		
Actual power	= 0.9501048		

H4: The interplay of follower exploration and exploitation (i.e., follower ambidexterity) positively predicts innovation outcomes, insofar that innovation is highest when both follower exploration and exploitation behaviors are high.

F tests - Linear multiple regression: Fixed model, R^2 increase

Analysis:	A priori: Compute required sample size		
Input:	Effect size f^2	=	0.05
α err prob	= 0.01		
Power (1-β err prob)	= 0.95		
Number of tested predictors	= 1		
Total number of predictors	= 4		
Output:	Noncentrality parameter λ	=	18.0000000
Critical F	= 6.7068004		
Numerator df	= 1		
Denominator df	= 355		
Total sample size	= 360		
Actual power	= 0.9502158		

Appendix B

Primary study	Code	Nationality	Industries/ Job Functions	Survey Administration	Sample Size	Effect size
Alghamdi (2018)	a	Saudi-Arabia	Public University	Paper-and-pencil (postal distribution)	147	H1: r =.41 H2: r =.45 H3: ΔR^2 =.04; f^2 =.06 H4: n.a.
Rosing & Zacher (2017) - Study 1	b.1	Australia	Range of occupations	Online	192	H1: n.a. H2: n.a. H3: n.a. H4: info not available from results*
Rosing & Zacher (2017) - Study 2	b.2	Germany	Creative industries	Online survey	156	H1: n.a. H2: n.a. H3: n.a. H4: info not available from results*
Zacher et al. (2016)	c	United States	Amazon Mturk participants	Online survey	388	H1: r = .42 H2: r =.21 H3: n.a. H4: ΔR^2 =.01; f^2 =.018
Zacher & Rosing (2015)	d	Australia	Architecture and Interiror Design	Paper-and-pencil (postal distribution)	33	H1: n.a. H2: n.a. H3: ΔR^2 =.12; f^2 =.21 H4: n.a.
Zacher & Wilden (2014)	e	unspecified	Broad range of jobs/occupations (administration officer, cleaner, customer service clerk, dance teacher, exercise physiologist, pharmacy assistant, and police officer etc.)	Online-survey	113	H1: n.a. H2: n.a. H3: info not available from results* H4: n.a.

Note. n.a. = not applicable; * = authors did not report ΔR^2

Appendix C

Task (identical in all conditions):

Background about the organization and the overall project:

We, that is, *[anonymized for review]* are planning a study in which participants will repeatedly (once per year) take part in a survey, over a period of 20 years.

Unfortunately, *[anonymized for review]* only has a restricted budget for compensating the participants which means that participant retention in the project needs to be ensured in other ways.

The main aim of this project is to track and examine how working in different jobs and how different job qualities impact people's lives over their lifespan.

The project especially focuses on capturing long-term effects (e.g., the effect of work over time on the respondents' identity, memory, cognition, and personality) and how the nature of work and workers changes. *[anonymized for review]* is planning to contact participants once a year to take part in questionnaire (a survey) that takes about 50 to 60 minutes to complete.

The findings are intended to support innovative solutions for the creation of high-quality work, as well as contribute to organizational and governmental policy on this topic.

Your task

We already have some material for our website which will be used as marketing material and for flyers.

For this task, we need help with this marketing material that will be used for recruitment purposes.

Your task is to:

- Enhance the marketability of the survey by **adding pictures** to the website material (using a license-free website, e.g., pixabay.com)
- Highlight important information within the text by **using different colors**.

Imagine that you will be working on this task with a supervisor. Your supervisor is named J.P.

In this simulation, J.P. will send you messages (emails) to provide you with guidance in accomplishing this task.

Appendix D

Draft of the flyer (that participants will receive at the beginning of the innovation task):

Website material

Home

Working Across the Career is a large, on-going, nationally representative longitudinal population study consisting of approximately 3,000 participants ranging from early to late adulthood. The main aim of the study is to examine work and its impact on people's lives over the life span. WALC especially focuses on capturing long-term effects (e.g., the effect of work over time on health, well-bing, identity, memory, learning and cognition) and how the nature of work and workers is changing. Our findings are intends to support the creation of good quality work for all, and contribute to organisational and governmental policy on this topic.

WALC study is a unique longitudinal study due to the following reasons:

It is the only known, large-scale longitudinal study to focus on work design, that is, how the work environments is organised and structured. This includes the characteristics of work such as autonomy, flexibility, colleague support, and ability to participate in decision-making. No other longitudinal study focuses so intensively on work and its relationships with causes and consequences.

Biological measures are obtained, including MRI, cortisol, and physical measures of health and well-being such as height, weight, and blood pressure.

Repeat assessment of participants across the lifespan allows investigation of the longitudinal, causal relationships between the nature of work, and aspects such as individual differences, identity, mental and physical health and well-being, memory, and cognition. We are thus able to study how age impact on these factors.

About us

The WALC study is run by the *[anonymized for review]* and was set up in partnership with *[anonymized for review]*.

For participants

The study

The Working Across the Career (WALS) study is the only longitudinal study of its type in the world, following individuals across the lifespan from early to late adulthood. It examines how work and the work environment changes across adult life, and the relationship between work and health and wellbeing, cognition and memory, identity, and individual differences such as personality.

The contribution of participants is to the continuation of our study. It is greatly appreciated and allows us to continue important research around creating good quality jobs for all, which is ever salient to us as individuals as well salient in the media.

$20 for participants

As a special thank you to each of our participants we will give each person $20. All you have to do is click on the survey below and fill it in. It should take no more than 30-40 minutes to complete.

Instant feedback report

We understand that individuals may be interested in results. Upon completion of the survey, we will send you an instant feedback report by email which details about you. For an example copy of this feedback report, please click here.

[Add a 'Participate now!' button to link to the online survey]

Take part in our substudies!

We are running two sub-studies which we would love you to participate in. These are:

Health and well-being
We are investigating how work impacts on mental and physical health and well-being. If you are willing to take part in this study please sign up here and one of us will be in contact with you shortly to arrange a suitable time for you to come in. All travel expenses will be paid.

Register for the Health and Well-being study

Memory and cognition
We are particularly interested in how work impacts on memory and cognition over the lifespan. If you would like to take part in this study, please sign up below and one of our team will be in contact shortly to arrange a suitable time for you to come in.

[Include a live counter at the bottom of the website indicating how many people have participated so far]

Confidentiality

It is very important to us that the information you provide is kept strictly confidential. We comply with the Privacy Act 1988, which means that your data is securely stored and protected, and only authorised people are allowed access to it.

Each participant is also asked for his or her consent before taking part in the survey and retains the right to withdraw at any time and / or ask for his or her data not to be used.

Recruitment

The WALC study is a ongoing longitudinol and representative population study. Participants were recruited using dual frame Random Digit Dialling (RDD), including both landlines and mobile phone numbers.

Appendix E

Condition 1

Opening leader behavior

Opening

Email 1
From: J.P.
Subject: Marketing and website material for 20-year longitudinal study

Hi, my name is J.P. and I want to strongly encourage you to play with different ideas and methods. That is, you will have plenty of possibilities for independent thinking and acting with me.

I think that it is okay to even deviate from the suggested instructions. There is no right or wrong way in doing this– it is much more important to learn from mistakes that we make. So I encourage you to take a risk, experiment as much as you can, and/or think about different ways and methods to make the most of this. Feel absolutely free to experiment with novel ideas and to try different working methods.

Opening

Email 2
From: J.P.
Subject: Re: Marketing and website material for 20-year longitudinal study

Hi, I just wanted to touch base with you and motivate you to engage in various ways of accomplishing this task.
Feel absolutely free to play and experiment with different ideas and methods. If you think that one working method is not particularly helpful, it is totally fine trying another one. I really think that it is worthwhile to take risks in a task like this.
I strongly encourage my workers to think and act independently. For task like this, there really are no right or wrong ideas. Do not be afraid to try new approaches. I think this is the only way that we can learn and become better.

Condition 2

Closing leader behavior

Closing

Email 1
From: J.P.
Subject: Marketing and website material for 20-year longitudinal study

Hi, my name is J.P. and I am here to provide you with specific instructions for this task. Overall, I think that the task instructions and formal requirements for this task are sufficiently clear and precise, so I would recommend that you stick closely to them.

I also want to mention that I regularly check the progress of those workers that I have to supervise. I do this to help you to stay on track. I hope this will help you in not deviating from the formal task instructions. I believe that close adherence to the rules and the formal task requirements is the best way to be successful here.

Closing

Email 2
From: J.P.
Subject: Re: Marketing and website material for 20-year longitudinal study

Hi, I just wanted to see how you are doing and use this opportunity to check your progress. If you have established a work routine, I would recommend sticking to it now. This ensures that you keep the overall goal in mind and allows you to finalize the task on time. It is good to stay focused now and concentrate on task completion.
It might also be worthwhile to check if basic task requirements are fulfilled. In the past, for example, some workers have uploaded the wrong document (that is, the original version) – so I encourage you to control carefully if you have uploaded the revised (not the original) document.

Condition 3

Ambidextrous leadership behavior

Opening

Email 1
From: J.P.
Subject: Marketing and website material for 20-year longitudinal study

Hi, my name is J.P. and I want to strongly encourage you to play with different ideas and methods. That is, you will have plenty of possibilities for independent thinking and acting with me.

I think that it is okay to even deviate from the suggested instructions. There is no right or wrong way in doing this– it is much more important to learn from mistakes that we make. So I encourage you to take a risk, experiment as much as you can, and/or think about different ways and methods to make the most of this. Feel absolutely free to experiment with novel ideas and to try different working methods.

Closing

Email 2
From: J.P.
Subject: Re: Marketing and website material for 20-year longitudinal study

Hi, I just wanted to see how you are doing and use this opportunity to check your progress. If you have established a work routine, I would recommend sticking to it now. This ensures that you keep the overall goal in mind and allows you to finalize the task on time. It is good to stay focused now and concentrate on task completion.
It might also be worthwhile to check if basic task requirements are fulfilled. In the past, for example, some workers have uploaded the wrong document (that is, the original version) – so I encourage you to control carefully if you have uploaded the revised (not the original) document.

Condition 4

Transformational leadership

Email 1
From: J.P.
Subject: Marketing and website material for 20-year longitudinal study

Hi, my name is J.P. and my positive vision for the future of this organization is to develop the best survey in the world. My goal is that our research should create significantly more insights compared to studies from other institutions – I want to reach this aim within the next five years.

I am very clear about my values for this organization and I am determined to practice what I preach. This project makes me very proud. I sincerely hope that I can inspire you to feel proud about it as well. I recognize and respect your contribution. I am extremely confident that you will show your best performance.

If you encounter any problems during this task, I am encouraging you to think about these problems in new ways (e.g., by questioning some of your assumptions).

My goal as a leader is to encourage your personal development and to pay attention to your individual needs.

Email 2:
From: J.P.
Subject: Re: Marketing and website material for 20-year longitudinal study

Hi, I will do my best to support you in your personal needs. I respect your work and encourage you to be proud of your accomplishments. I have a very clear sense of where I see this organization in the future.
I know that you will give your very best performance in this task and I will acknowledge your accomplishments when I see outstanding work.

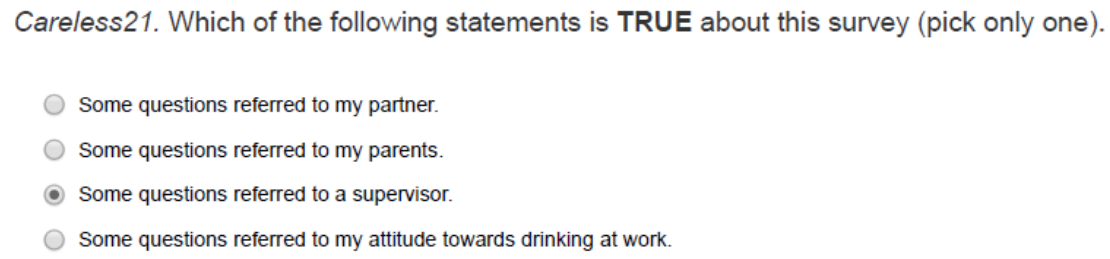
Appendix F

To control for careless responding (or for participants who did not pay sufficient attention), we included four attention check measures (at the end of the survey) that allowed us to identify participants that did not pay attention to the contents of the experiment.

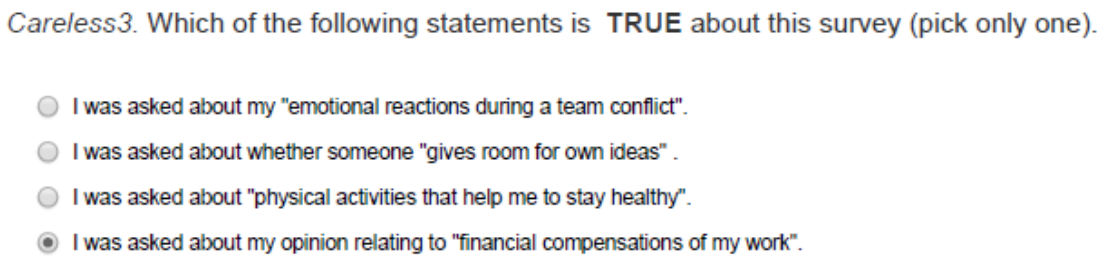
Attention check (Step 1)

In the first step, we used two comprehension measures in which we asked participants to correctly identify previous elements of the survey ("Which of the following statements is TRUE about this survey (pick only one)").

For the first measure, participants had to select one out of the following four options: "Some questions referred to my partner." (incorrect), "Some questions referred to my parents." (incorrect), "Some questions referred to a supervisor." (correct), and "Some questions referred to my attitude towards drinking at work." (incorrect).



For the second measure, participants had to select one out of the following four options: "I was asked about my 'emotional reactions during a team conflict' (incorrect), "I was asked about whether someone 'gives room for own ideas'." (correct), "I was asked about 'physical activities that help me to stay healthy' (incorrect), and "I was asked about my opinion relating to 'financial compensations of my work'." (incorrect).



Attention check (Step 2)

In a second step, participants had to correctly identify the email (i.e., the experimental condition) that they were presented with at the beginning of the study ("Please indicate which of following emails resembles most accurately the first/second email message that you received from your supervisor?").

Since all participants had received two emails, they also had to correctly identify the order of the emails [i.e., first they had to identify the first email correctly (attention check measure 3); then they had to correctly identify the second email (attention check measure 4)]. That is, for the third attention check measure (correct identification of email 1), participants had to select one out of three options (opening email 1, closing email 1, or transformational email 1; the full text that was presented to participants is given in Appendix E). For the fourth attention check measure (correct identification of email 2), participants had to select one out of three options (opening email 2, closing email 2, or transformational email 2; see full text for each email is given in Appendix E). To pass the third and fourth attention check measures, participants had to select the correct options (that is, picking the email that was in line with the manipulation they had received previously). Participants were only included in our final analysis if they passed all four manipulation checks.

AttentionCheck1. Please indicate which of following emails resembles most accurately the **first email message** that you received from your supervisor?

◉ E-mail 1

From: J.P.
Subject: Re: MY TASK

Hi Jeremiah, my name is J.P. and I want to strongly encourage you to play with different ideas and methods. That is, you will have plenty of possibilities for independent thinking and acting with me.

I think that it is okay to even deviate from the suggested instructions. There is no right or wrong way in doing this– it is much more important to learn from mistakes that we make. So I encourage you to take a risk, experiment as much as you can, and/or think about different ways and methods to make the most of this. Feel absolutely free to experiment with novel ideas and to try different working methods.

○ E-mail 1

From: J.P.
Subject: Re: MY TASK

Hi Jeremiah, my name is J.P. and I am here to provide you with specific instructions for this task. Overall, I think that the task instructions and formal requirements for this task are sufficiently clear and precise, so I would recommend that you stick closely to them.

I also want to mention that I regularly check the progress of those workers that I have to supervise. I do this to help you to stay on track. I hope this will help you in not deviating from the formal task instructions. I believe that close adherence to the rules and the formal task requirements is the best way to be successful here.

○ E-mail 1

From: J.P.
Subject: Re: MY TASK

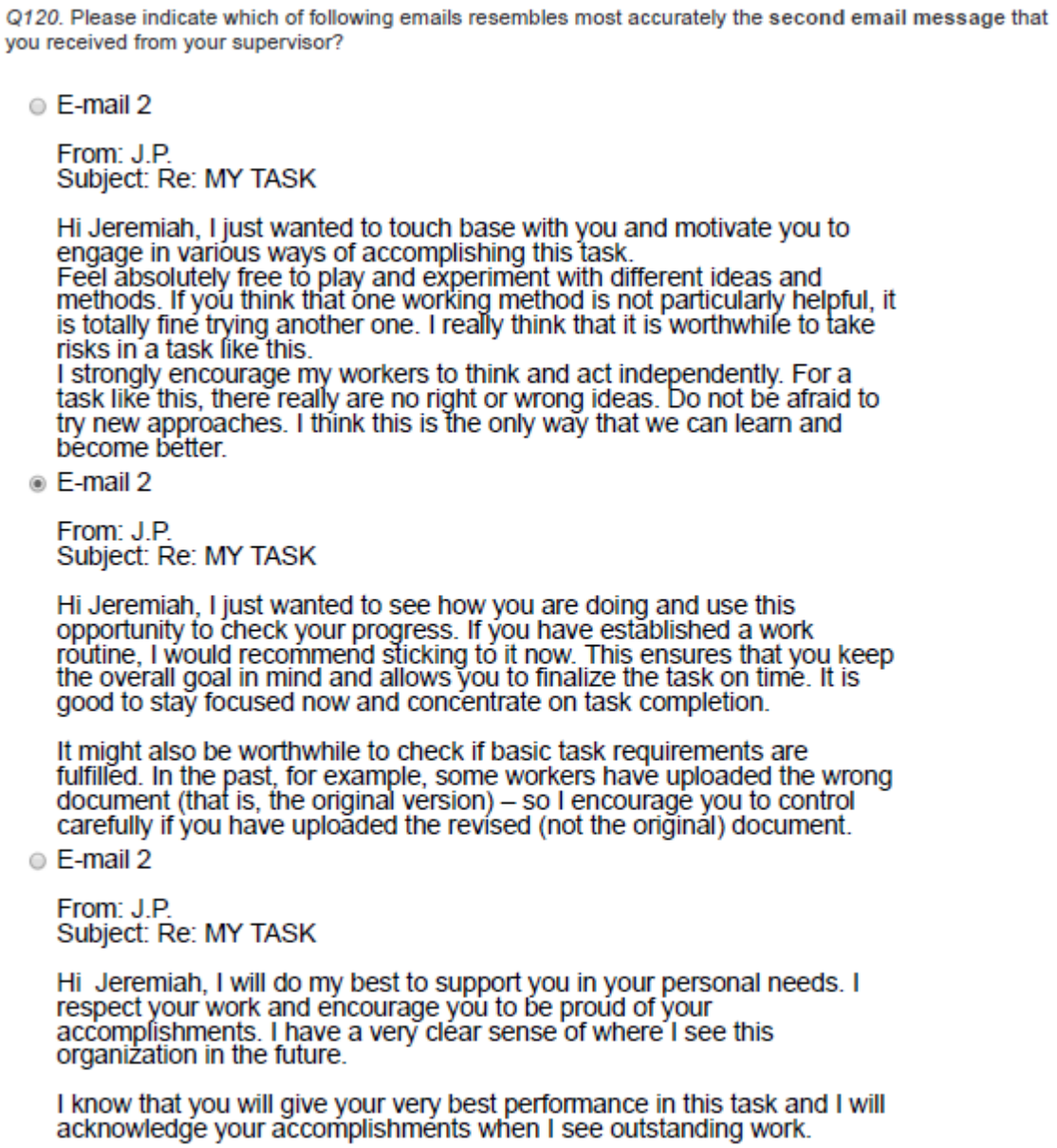
Hi Jeremiah, my name is J.P. and my positive vision for the future of this organization is to develop the best survey in the world. My goal is that our research should create significantly more insights compared to studies from other institutions – I want to reach this aim within the next five years.

I am very clear about my values for this organization and I am determined to practice what I preach. This project makes me very proud. I sincerely hope that I can inspire you to feel proud about it as well. I recognize and respect your contribution. I am extremely confident that you will show your best performance.

If you encounter any problems during this task, I am encouraging you to think about these problems in new ways (e.g., by questioning some of your assumptions).

My goal as a leader is to encourage your personal development and to pay attention to your individual needs.

Note: Participants were given the possibility to indicate their first name (before the experiment) so "that their supervisor J.P. knew how to address them". We used a piped text function in Qualtrics which allowed us to customize the emails in a way that participants were addressed by their first name in each email.



Q120. Please indicate which of following emails resembles most accurately the **second email message** that you received from your supervisor?

○ E-mail 2

From: J.P.
Subject: Re: MY TASK

Hi Jeremiah, I just wanted to touch base with you and motivate you to engage in various ways of accomplishing this task.
Feel absolutely free to play and experiment with different ideas and methods. If you think that one working method is not particularly helpful, it is totally fine trying another one. I really think that it is worthwhile to take risks in a task like this.
I strongly encourage my workers to think and act independently. For a task like this, there really are no right or wrong ideas. Do not be afraid to try new approaches. I think this is the only way that we can learn and become better.

◉ E-mail 2

From: J.P.
Subject: Re: MY TASK

Hi Jeremiah, I just wanted to see how you are doing and use this opportunity to check your progress. If you have established a work routine, I would recommend sticking to it now. This ensures that you keep the overall goal in mind and allows you to finalize the task on time. It is good to stay focused now and concentrate on task completion.

It might also be worthwhile to check if basic task requirements are fulfilled. In the past, for example, some workers have uploaded the wrong document (that is, the original version) – so I encourage you to control carefully if you have uploaded the revised (not the original) document.

○ E-mail 2

From: J.P.
Subject: Re: MY TASK

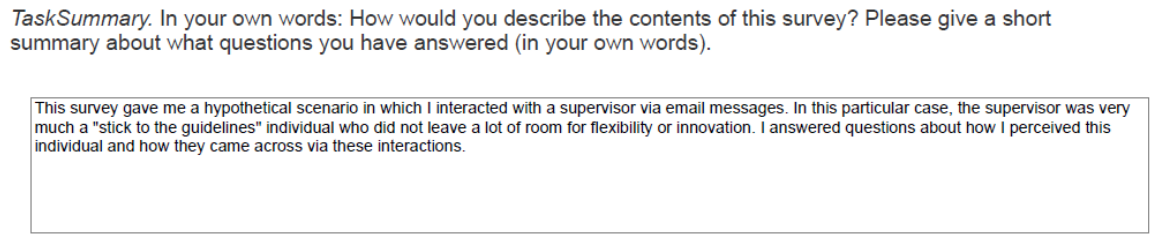
Hi Jeremiah, I will do my best to support you in your personal needs. I respect your work and encourage you to be proud of your accomplishments. I have a very clear sense of where I see this organization in the future.

I know that you will give your very best performance in this task and I will acknowledge your accomplishments when I see outstanding work.

Note: Participants were given the possibility to indicate their first name (before the experiment) so "that their supervisor J.P. knew how to address them". We used a piped text function in Qualtrics which allowed us to customize the emails in a way that participants were addressed by their first name in each email.

Attention check (Step 3)

Finally, we added an open question prompting a task summary at the end of the survey ("In your own words: How would you describe the contents of this survey? Please give a short summary about what questions you have answered") that allowed us to qualitatively check if participants paid attention



We excluded 77 participants (50%) who failed attention check measures (for this sample, we also checked the open text responses that participants provided to the supervisor email. This provided further validation that participants who failed the attention checks did not pay attention/were careless with the overall task (e.g., by copying random text in their response email; responding "great" to the task summary question instead of giving a description of the task). In contrast, participants who passed attention checks ($N = 75$) provided strong evidence that they were highly engaged in the task (participants directly responded to elements of the supervisor email; e.g., (after opening email): "Okay I agree with learning from mistakes. It is the best way to come up with new ideas. Do you have any other directions for me before I start? I'll get started right away.").

While a rejection rate of 50% seems relatively high, a recent review on Mturk and online studies has argued that this rejection rate is still within the range (albeit on the higher end) of rates reported in previous research (Thomas & Clifford, 2017, indicated a range of 2% to 52% for Mturk studies and 6 to 46% for lab studies). Most importantly, the authors noted that higher rates "tend to result from multiple, objectively verifiable comprehension checks about experimental materials", p. 190, which reflects the nature of our screening items). Furthermore, rejection (lack of attention, carefulness or comprehension) did not result in any biased over-sampling of one of the four experimental conditions ($\chi^2(3) = .50$, $p = .92$)

Appendix G

Instruction for subject matter experts to assess participants' workplace innovation[1]:

Compared to other solutions that you have seen for this task, this task outcome…				
Is a usual and conventional solution (1)	Is a solution that features only a few novel aspects (2)	Has a balance between conventional, usual, and novel aspects (3)	Is a solution that features mostly novel aspects (4)	Is completely novel and does not at all rely on conventional solutions (5)
Is not creative at all (1)	Is somewhat uncreative (2)	Is neither creative nor uncreative (3)	Is somewhat creative (4)	Is very creative (5)
Is insufficient in terms of quality standards (1)	Falls somewhat below the quality standards (2)	Is acceptable in terms of quality standards (3)	Meets the quality standards very well (4)	Exceeds the quality standards (5)
Cannot be applied in the 'real world' (i.e., in professional settings) (1)	Can be applied in the 'real world' with major modifications (2)	Can be applied in the 'real world' with some modifications (3)	Can be applied in the 'real world' with little modifications (4)	Can readily be applied in the 'real world' (5)

[1] Items based on Rosing, Bledow, Frese, Baytalskaya, Lascano, and Farr (2018)

Appendix H

1. Self-report measure used in PILOT STUDY (without leader manipulations)

Measurement of innovative behavior

Source: Zacher et al. (2016)

Instructions:
Imagine that the materials that you have just revised and submitted were to be reviewed by a supervisor or boss.
'In your honest opinion, how do you think your supervisor or boss would rate you in this task based on the descriptions below?'

Items:
1. Coming up with new ideas.
2. Working to implement new ideas.
3. Findings improved ways to do things.
4. Creating better processes and routines.

Answer scale:
1 = needs much improvement, 2 = needs some improvement, 3 = satisfactory, 4 = good, 5 = excellent

2. Alternative measures to be used as INSTRUMENTAL VARIABLES:

Measurement of Positive (1.-5.) and Negative Trait Affect (6.-10.)

Source:
Original: Mackinnon, Jorm, Christensen, Korten, Jacomb, & Rodgers (1999)
*Adaptation: (of the answer scales): Zacher et al. (2016)

Instructions:
This scale consists of a number of words that describe different feelings and emotions. Read each item and then list the number from the scale below next to each word.

Indicate to what **extent you feel this way in general**:

1. Inspired
2. Alert
3. Excited
4. Enthusiastic
5. Determined

6. Afraid
7. Upset
8. Nervous
9. Scared
10. Distressed

Answer scale:

1 = very slightly or not at all, 5 = extremely*

Measurement of Conscientiousness (HEXACO):

Source: Ashton & Lee (2009)

Instructions:
On the following pages, you will find a series of statements about you. Please read each statement and decide how much you agree or disagree with that statement. Then indicate your response using the following scale:

1. I plan ahead and organize things, to avoid scrambling at the last minute.
2. I often push myself very hard when trying to achieve a goal.
3. When working on something, I don't pay much attention to small details. (reverse-scored)
4. I make decisions based on the feeling of the moment rather than on careful thought. (reverse-scored)
5. When working, I sometimes have difficulties due to being disorganized. (reverse-scored)
6. I do only the minimum amount of work needed to get by. (reverse-scored)
7. I always try to be accurate in my work, even at the expense of time.
8. I make a lot of mistakes because I don't think before I act. (reverse-scored)
9. People often call me a perfectionist.
10. I prefer to do whatever comes to mind, rather than stick to a plan. (reverse-scored)

Measurement of Openness to experience (HEXACO):

1. I would be quite bored by a visit to an art gallery. (reverse-scored)
2. I'm interested in learning about the history and politics of other countries.
3. I would enjoy creating a work of art, such as a novel, a song, or a painting.
4. I think that paying attention to radical ideas is a waste of time. (reverse-scored)
5. If I had the opportunity, I would like to attend a classical music concert.
6. I've never really enjoyed looking through an encyclopedia. (reverse-scored)
7. People have often told me that I have a good imagination.
8. I like people who have unconventional views.
9. I don't think of myself as the artistic or creative type. (reverse-scored)
10. I find it boring to discuss philosophy. (reverse-scored)

Answer scale:
1 = strongly disagree, 5 = strongly agree

Appendix I.

Example 1 of task outcome (submitted by participant).
Low innovation rating

[Innovation rating from rater (1) was *M* = 1, innovation rating from rater (2): M = 1.25; both raters used the 4 item innovation rating instrument with a 5-point Likert scale from Rosing et al., 2018)

Website material

Home

Working Across the Career is a large, on-going, nationally representative longitudinal population study consisting of approximately 3,000 participants ranging from early to late adulthood. The main aim of the study is to examine work and its impact on people's lives over the life span. WALC especially focuses on capturing long-term effects (e.g., the effect of work over time on health, well-being, identity, memory, learning and cognition) and how the nature of work and workers is changing. Our findings are intends to support the creation of good quality work for all, and contribute to organisational and governmental policy on this topic.

WALC study is a unique longitudinal study due to the following reasons:

It is the only known, large-scale longitudinal study to focus on work design, that is, how the work environments is organised and structured. This includes the characteristics of work such as autonomy, flexibility, colleague support, and ability to participate in decision-making. No other longitudinal study focuses so intensively on work and its relationships with causes and consequences.

Biological measures are obtained, including MRI, cortisol, and physical measures of health and well-being such as height, weight, and blood pressure.

Repeat assessment of participants across the lifespan allows investigation of the longitudinal, causal relationships between the nature of work, and aspects such as individual differences, identity, mental and physical health and well-being, memory, and cognition. We are thus able to study how age impact on these factors.

For participants

The study

The Working Across the Career (WALS) study is the only longitudinal study of its type in the world, following individuals across the lifespan from early to late adulthood. It examines how work and the work environment changes across adult life, and the relationship between work and health and wellbeing, cognition and memory, identity, and individual differences such as personality.

The contribution of participants is to the continuation of our study. It is greatly appreciated and allows us to continue important research around creating good quality jobs for all, which is ever salient to us as individuals as well salient in the media.

$20 for participants

As a special thank you to each of our participants we will give each person $20 as compensation. All you have to do is click on the survey below and fill it in. It should take no more than 30-40 minutes to complete.

Instant feedback report

We understand that individuals may be interested in results. Upon completion of the survey, we will send you an instant feedback report by email which details about you. For an example copy of this feedback report, please click here.

[Add a 'Participate now!' button to link to the online survey]

Take part in our sub-studies!

We are running two sub-studies which we would love you to participate in. These are:

Health and well-being
We are investigating how work impacts on mental and physical health and well-being. If you are willing to take part in this study please sign up here and one of us will be in contact with you shortly to arrange a suitable time for you to come in. All travel expenses will be paid.

Register for the Health and Well-being study

Memory and cognition
We are particularly interested in how work impacts on memory and cognition over the lifespan. If you would like to take part in this study, please sign up below and one of our team will be in contact shortly to arrange a suitable time for you to come in.

[Include a live counter at the bottom of the website indicating how many people have participated so far]

Confidentiality

It is very important to us that the information you provide is kept strictly confidential. We comply with the Privacy Act 1988, which means that your data is securely stored and protected, and only authorised people are allowed access to it.

Each participant is also asked for his or her consent before taking part in the survey and retains the right to withdraw at any time and / or ask for his or her data not to be used.

Recruitment

The WALC study is a ongoing longitudinal and representative population study. Participants were recruited using dual frame Random Digit Dialling (RDD), including both landlines and mobile phone numbers.

Example 2 of task outcome (submitted by participant).
High innovation rating
[Innovation rating from rater (1) was *M* = 3.25, innovation rating from rater (2): M = 3.5; both raters used the 4 item innovation rating instrument with a 5-point Likert scale from Rosing et al., 2018)

Working Across the Career Study

Working Across the Career is a large, on-going, nationally representative longitudinal population study consisting of approximately 3,000 participants ranging from early to late adulthood. The main aim of the study is to examine work and its impact on people's lives over the studys lifespan.

WALC especially focuses on capturing long-term effects (e.g., the effect of work across time on health, well-being, identity, memory, learning and cognition) and how the nature of work and workers is changing. Our findings are intended to support the creation of good quality work for all, and contribute to organisational and governmental policy on this topic.

WALC Study is a Unique Longitudinal Study

It is the only known, large-scale longitudinal study to focus on work design, that is, how the work environments is organised and structured. This includes the characteristics of work such as autonomy, flexibility, colleague support, and ability to participate in decision-making.

No other longitudinal study focuses so intensively on work and its relationships with causes and consequences!

Biological measures are obtained, including MRI, cortisol, and physical measures of health and well-being such as height, weight, and blood pressure.

Repeat assessment of participants across the lifespan allows investigation of the longitudinal, causal relationships between the nature of work, and aspects such as individual differences, identity, mental and physical health and well-being, memory, and cognition. We are thus able to study the impact of age on these factors.

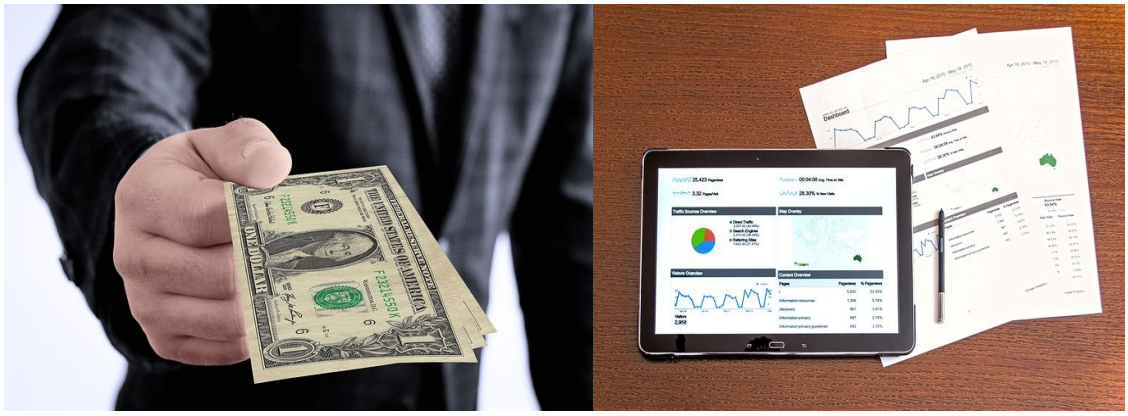
About the Study & Benefits

The Working Across the Career (WALS) study is the only longitudinal study of its type in the world, following individuals across the lifespan from early to late adulthood. It examines how work and the work environment changes across adult life, and the relationship between work and health and wellbeing, cognition and memory, identity, and individual differences such as personality.

The contribution of participants is to the continuation of our study. It is greatly appreciated and allows us to continue important research around creating good quality jobs for all, which is ever salient to us as individuals as well salient in the media.

$20 for participants

As a special thank you to each of our participants we will give each person $20. All you have to do is click on the survey below and fill it in. It should take no more than **30-40 minutes to complete**.



Instant feedback report

We understand that individuals may be interested in results. Upon completion of the survey, we will send you an instant feedback report by email which details about you. For an example copy of this feedback

report, please *click* *here*.

[CLICK TO PARTICIPATE NOW]

Take Part in our Substudies

We are running two sub-studies which we would love you to participate in.

Health and Well-being

We are investigating how work impacts on mental and physical health and well-being. If you are willing to take part in this study please sign up here and one of us will be in contact with you shortly to arrange a suitable time for you to come in. All travel expenses will be paid.

[Click HERE to Register for the Health and Well-being study]

Memory and Cognition

We are particularly interested in how work impacts on memory and cognition over the lifespan. If you would like to take part in this study, please sign up below and one of our team will be in contact shortly to arrange a suitable time for you to come in.

[Click HERE to Register for the Memory & Cognition study]

[Include a live counter at the bottom of the website indicating how many people have participated so far]

Further Information

Confidentiality

It is very important to us that the information you provide is kept strictly confidential. We comply with the Privacy Act 1988, which means that your data is securely stored and protected, and only authorised people are allowed access to it.

Each participant is also asked for his or her consent before taking part in the survey and retains the right to withdraw at any time and / or ask for his or her data not to be used.

Recruitment

The WALC study is a ongoing longitudinol and representative population study. Participants were recruited using dual frame Random Digit Dialling (RDD), including both landlines and mobile phone numbers.

Appendix J

Overview of analytical steps that will be performed to test the hypotheses for Study 1 and Study 2.

Nr.	Step	Analytical Approach	SPSS Syntax / Mplus Syntax	Key decisions for results / Ex-ante criteria for successful replication
1	**Preparation/ cleaning of data:** Exclusion of participants that do not pass attention and comprehension checks.	a. Dummy coding of attention and comprehension checks items (1 = passed attention check, 0 = did not pass attention check) b. Create a composite score for all attention check items and use it as a filter for the final analysis	**SPSS** a. Example for one of the attention check items. RECODE Email1Identified (0=1) (Else =0) INTO Filter_CorrectEmail1 . EXECUTE. VALUE LABELS / Filter_CorrectEmail 0 "Did not pass the attention check" 1 "Passed the attention check". CACHE. EXECUTE. b. Compute Filter_AttentiveParticipant = Filter_CorrectEmail1 * Filter_CorrectEmail2 * CarelessResponding *AttenitonCheck1 * AttentionCheck2. EXECUTE. VALUE LABELS / Filter_AttentiveParticipant 0 "Did not pass one of the multiple attention checks" 1 "Passed all of the multiple attention checks". CACHE. EXECUTE.	If exclusion criteria minimizes the required sample size, we will continue to collect data until we reach the required sample size of $N = 400$
2	**Check construct validity of survey-based measures:**		**MPLUS** usevar = Explor1 Explor2 Explor3 Explor4 Explor5 Explor6 Explor7 Exploit1 Exploit2 Exploit3 Exploit4 Exploit5 Exploit6**;**	If the χ^2 rejects the model, we will check modification indices for sources of potential misfit and discuss the sources of potential

	For exploration and exploitation, we will evaluate the measurement fit for the two-dimensional factor structure		missing = all (999);
			analysis: type = general;
			model:
			ExplorFac by Explor1@1 Explor2 Explor3 Explor4 Explor5 Explor6 Explor7;
			
			ExploiFac by Exploit1@1 Exploit2 Exploit3 Exploit4 Exploit5 Exploit6;
			! Specification of measurement model for exploration and exploitation
			
			output: sampstat standardized stdyx;
				misfit within the discussion section of our study.

2	**Check psychometrics of survey-based measures:** Calculate reliability for all multi-item measures	Calculation of Cronbach's alpha for each multi-item, summarizing multi-item scales into one measure	**SPSS** Example for exploration behavior: *Estimation of cronbach's alpha (reliability). RELIABILITY /VARIABLES= Explor1 Explor2 Explor3 Explor4 Explor5 Explor6 Explor7 /SCALE('ALL VARIABLES') ALL /MODEL=ALPHA /STATISTICS=DESCRIPTIVE SCALE CORR /SUMMARY=TOTAL MEANS VARIANCE. EXECUTE. *Aggregation of items into one scale score. COMPUTE Explor_7itemMean =Mean(Explor1 to Explor7). EXECUTE.	Report Cronbach's alpha in method sections
3	Assessing the **rater agreement** for innovation outcome	Calculation of intra-class correlation for a subset of double-coded outcomes	**SPSS** *Computation of mean using the four items from on Rosing, Bledow, Frese, Baytalskaya, Lascano, and Farr (2018) for Rater 1 and Rater 2. COMPUTE Inno_mean_Rat1 =Mean(Inno1_rat1, Inno2_rat1, Inno3_rat1, Inno4_rat1). EXECUTE. COMPUTE Inno_mean_Rat2 =Mean(Inno1_rat2, Inno2_rat2, Inno3_rat2, Inno4_rat2). EXECUTE.	In case agreement is ICC ≤ .60, the raters discuss the misalignments for those outcomes in which ratings differed considerably and come to an agreement in consultation with the

			*Computation of Intra-Class Correlation between two raters. RELIABILITY /VARIABLES= Inno_mean_Rat1 Inno_mean_Rat2 /SCALE('ALL VARIABLES') ALL /MODEL=ALPHA /STATISTICS=DESCRIPTIVE SCALE CORR /SUMMARY=MEANS /ICC=MODEL(MIXED) TYPE(ABSOLUTE) CIN=95 TESTVAL=0. ** Compute Mean Score for both Raters. COMPUTE Inno_RaterMean =Mean(Inno_mean_Rat1, Inno_mean_Rat2). EXECUTE.	first and second authors.
4	Testing H1	Calculation of one-factorial ANOVA and Contrasts between conditions.	**SPSS.** *Recoding experimental conditions. AUTORECODE VARIABLES=Condition /INTO Condition_num /PRINT. VALUE LABELS /Condition_num 1 "Email: OPENING" 2 "Email: CLOSING" 3 "Email: AMBIDEXTROUS" 4 "Email:TRANSFORMATIONAL". CACHE. EXECUTE. *Test if EXPLORATION is significantly different between different leadership conditions. A significant F-value will only tell us that exploration differs between conditions (but not which leadership condition has the strongest effect on exploration or which specific leadership conditions differ from each other). ONEWAY Explor_7itemMean BY Condition_Num /STATISTICS DESCRIPTIVES EFFECTS HOMOGENEITY /PLOT MEANS	**Ex-ante criteria for successful replication** H1 is supported when a. Opening condition > Closing condition ($p < .05$) b. Opening condition > Transformational condition ($p < .05$)

		/MISSING ANALYSIS. ** Testing contrasts to evaluate specific differences between two experimental conditions *Do participants show significantly more "Exploration" in the "OPENING condition" (contrasting against CLOSING). T-TEST GROUPS=Condition_num(1 2) /MISSING=ANALYSIS /VARIABLES= Explor_7itemMean /CRITERIA=CI(.95). *Do participants show significantly more "Exploration" in the "OPENING condition" (contrasting against TRANSFORMATIONAL). T-TEST GROUPS=Condition_num(1 4) /MISSING=ANALYSIS /VARIABLES= Explor_7itemMean /CRITERIA=CI(.95).	
Testing H2	Calculation of one-factorial ANOVA and Contrasts between conditions.	**SPSS.** *Test if EXPLOITATION is significantly different between different leadership conditions. A significant F-value will only tell us that exploitation differs between conditions (but not which leadership condition has the strongest effect on exploration or which specific leadership conditions differ from each other. ONEWAY Exploit_6itemMean BY Condition_Num /STATISTICS DESCRIPTIVES EFFECTS HOMOGENEITY /PLOT MEANS /MISSING ANALYSIS. ** Test contrasts to evaluate specific differences between two experimental conditions *E.g. do participants show significantly more "Exploitation" in the "CLOSING condition" (contrasting against Opening). T-TEST GROUPS=Condition_num(1 2)	**Ex-ante criteria for successful replication** H2 is supported when a. Closing condition > Opening condition ($p < .05$) b. Closing condition > Transformational condition ($p < .05$)

		/MISSING=ANALYSIS /VARIABLES= Exploit_6itemMean /CRITERIA=CI(.95).	
Testing H3	Calculation of one-factorial ANOVA and Contrasts between conditions.	**SPSS.** *Test if INNOVATION is significantly different between different leadership conditions. A significant F-value will only tell us that INNOVATION differs between conditions (but not which leadership condition has the strongest effect on exploration or which specific leadership conditions differ from each other. ONEWAY Inno_RaterMean BY Condition_Num /STATISTICS DESCRIPTIVES EFFECTS HOMOGENEITY /PLOT MEANS /MISSING ANALYSIS. ** Test contrasts to evaluate specific differences between two experimental conditions *E.g. do participants have significantly higher scores for "Innovation" in the "AMBIDEXTROUS condition" (contrasting against TRANSFORMATIONAL). T-TEST GROUPS=Condition_num(3 4) /MISSING=ANALYSIS /VARIABLES= Inno_RaterMean /CRITERIA=CI(.95).	**Ex-ante criteria for successful replication** H3 is supported when a. Ambidextrous condition > Closing condition ($p < .05$) b. Ambidextrous condition > Opening condition ($p < .05$) c. Ambidextrous condition > Transformational condition ($p < .05$)
Testing H4	2SLS procedure (to purge endogenous variables from endogeneity) and multi-level regressions to test the interaction effect of follower ambidexterity	**SPSS.** ** Preparation of instrumental variables: Re-coding of experimental variable (= conditions) into binary dummy codes for each condition RECODE Condition_num (1=1) (ELSE=0) INTO Opening_dichotomous. VARIABLE LABELS Opening_dichotomous 'Opening condition vs. other'. EXECUTE. VALUE LABELS / Opening_dichotomous 1 "OPENING Leadership"	**Ex-ante criteria for successful replication** H4 is supported when Parameter estimate in the 2SLS is significant ($p < .05$)

```
2          "OTHER Leadership conditions (not Opening)".
CACHE.
EXECUTE.

RECODE Condition_num (2=1) (ELSE=0) INTO Closing_dichotomous.
VARIABLE LABELS  Closing_dichotomous 'Closing condition vs. other'.
EXECUTE.
VALUE LABELS
          / Closing_dichotomous
1          "CLOSING Leadership"
2          "OTHER Leadership conditions (not Closing)".
CACHE.
EXECUTE.

**Calculation of the follower ambidexterity variable.
COMPUTE Follower_Ambi = Exploit_6itemMean  * Explor_7itemMean  .
EXECUTE.
```

STATA.

```
************************
*2 stage-least square procedure.
************************
!** Running 2SLS (2-stage least square, cf., Antonakis et al., 2010, p. 1104).

ivreg Inno_RaterMean  (Follower_Ambi = Closing_dichotomous
Opening_dichotomous)

!** the command “ivreg” carries out the 2SLS (Instrumental variable regression !** in
STATA), the outcome variable Y is “Inno_RaterMean” which is predicted by the
endogenous variable “Follower_Ambi” (this endogenous predictor is instrumented on
the two conditions “Closing_dichotomous” and “Opening_dichotomous”


** We will first evaluate if the instruments are strong enough (F-value > 10 for
regression, cf., Staiger & Stock, 1997). If this is not the case, we will use other
instrumental variable (i.e., personality, trait positive affect)
```