What You See is What You Get?
Enhancing Methodological Transparency in Organizational Behavior Research

Abstract

The goal of our paper is to improve the inferential reproducibility and, consequently, the credibility and trustworthiness of organizational behavior and management research by providing evidence-based best practices on how to enhance methodological transparency. We review the literature on evidence-based best practices regarding how to enhance methodological transparency when conducting organizational behavior and management research. We distill the large, fragmented, and often-technical literature on how to enhance transparency into 44 recommendations from 96 review articles published in 28 substantive and methodological journals in management, business, sociology, and psychology. These specific and actionable recommendations are described in detail and summarized in six separate tables in our full-length manuscript. Our recommendations provide guidance on how to increase transparency at each stage of the research process: (1) theory (e.g., are a priori and post hoc hypotheses clearly designated as such? what are the epistemological assumptions?); (2) design (e.g., what is the sampling strategy? what is the rationale for including each control variable?); (3) measurement (e.g., are existing scales altered in any way? what is the justification for score aggregation?); (4) analysis (e.g., how are outliers handled? are data and syntax made available?); and (5) reporting of results (e.g., what is the justification for handling missing data? what is the total number of statistical tests conducted?). Our recommendations provide a resource for doctoral education, training, researchers, journal editors, reviewers, and journals and professional associations interested in enhancing methodological transparency in conducting research and the manuscript submission and review process.

Summary

Organizational behavior, management in general, and many other fields are currently debating the credibility, trustworthiness, and usefulness of the scholarly knowledge that is produced (Davis, 2015; George, 2014). It is worrisome that from 2005 through 2015, 129 articles have been retracted from business and management journals, and the number of retractions has increased tenfold from 2012 through 2015 (Karabag & Berggren, 2016). In addition, 25% to 50% of published articles in management and other fields have inconsistencies or errors (Goldfarb & King, 2016; Nuijten, Hartgerink, Assen, Epskamp, & Wicherts 2016; Wicherts, Bakker, & Molenaar, 2011). Overall, there is a proliferation of evidence indicating substantial reasons to doubt the veracity and, justifiably, the conclusions and implications of scholarly work (e.g., Banks, Rogelberg, Woznyj, Landis, & Rupp, 2016; Schwab & Starbuck, 2017) because researchers are often unable to reproduce published results (Bakker, van Dijk, & Wicherts, 2012; Bergh, Sharp, Aguinis, & Li, in press; Bergh, Sharp, & Li, 2017; Cortina, Green, Keeler, & Vandenberg, 2017). Regardless of whether this lack of reproducibility is a more recent phenomenon, or one that has existed for a long time but has only recently gained prominence, it seems that we have reached a tipping point such that there is an urgency to understand this phenomenon and find solutions to address it.

Our position is that concerns about lack of reproducibility are not entirely surprising considering the relative lack of methodological transparency about the process of conducting

empirical research that eventually leads to a published article (Banks, O'Boyle, et al., 2016; Bedeian, Taylor, & Miller, 2010; John, Loewenstein, & Prelec, 2012; O'Boyle, Banks, & Gonzalez-Mulé, 2017; Schwab & Starbuck, 2017; Simmons, Nelson, & Simonsohn, 2011; Wicherts et al., 2011; Wigboldus & Dotsch, 2016). We define methodological transparency *as the degree of detail and disclosure about the specific steps, decisions, and judgment calls made during a scientific study*. Based on this definition, we conceptualize transparency as a continuum—a matter of degree—and not as a dichotomous variable (i.e., transparency is present or absent). Clearly, researchers make numerous choices, judgment calls, and decisions during the process of conceptualizing and designing studies, as well as collecting data, analyzing them, and reporting results. The more explicit, open, and thorough researchers are about disclosing each of these choices, judgment calls, and decisions, the greater the degree of methodological transparency.

Low methodological transparency has a detrimental impact on the credibility and trustworthiness of research results because it precludes *inferential reproducibility*. Inferential reproducibility is the ability of others to draw similar conclusions to those reached by the original authors regarding a study's results (Goodman, Fanelli, & Ioannidis, 2016). Note that this is different from *results reproducibility*, which is the ability of others to obtain the same results using the same data as in the original study. From a measurement perspective, results reproducibility is conceptually analogous to reliability because it is about consistency. Specifically, do researchers other than those who authored a study find the same (i.e., consistent) results as reported in the original paper? Inferential reproducibility is conceptually analogous to validity because it is about making similar inferences based on the results. Specifically, do researchers other than those who authored a study reach similar conclusions about relations between variables as described in the original study? Results reproducibility (i.e., reliability) is a necessary but insufficient precondition for inferential reproducibility (i.e., validity). In other words, if we cannot obtain the same results as in the published study using the same data, inferences are clearly going to be different. But, it is possible to reproduce results (i.e., high reliability) but not inferences (i.e., low validity).

Inferential reproducibility (i.e., validity or the nature and direction of relations between variables) is the critical issue in terms of building and testing theories and the credibility of the knowledge that is produced, while results reproducibility (i.e., reliability or consistency) is a means to an end. For example, assume that a team of researchers uses archival data and publishes an article reporting a test of a model including five variables with satisfactory fit statistics. Then, a separate team of researchers uses the same dataset with the same five variables and is able to reproduce the exact same results (i.e., high reliability). This is a situation with a high degree of results reproducibility. Now, assume that, unbeknownst to the second team, the first team of researchers had tested 50 different configurations of variables and in the end, they found and reported the one configuration of the five variables that resulted in the best possible fit statistics. Obviously, testing so many configurations maximized capitalization on chance, and the good fit of the final model is more likely due to chance rather than substantive relations (Aguinis, Cascio, & Ramani, in press). Enhancing methodological transparency by disclosing that 50 different configurations of variables were tested until the final set was found would not affect results reproducibility, but it would certainly change inferential reproducibility. That is, the second team of researchers would reach very different inferences from the same results because the good fit of the model would be attributed more to chance than to the existence of substantive relations between variables.

Many articles published in organizational behavior and other fields represent situations similar to the example described above regarding the description of a model including five variables: We *simply do not know whether what we see is what we get*. Most things seem just right: measures are valid and have good psychometric qualities, hypotheses described in the Introduction are mostly supported by results, statistical assumptions are not violated (or not mentioned), the "storyline" is usually neat and straightforward, and everything looks in place. But, unbeknownst to readers, many researchers have engaged in various trial-and-error practices (e.g., revising, dropping, and adding scale items), opaque choices (e.g., including or excluding control variables), and other decisions (e.g., altering scales, retroactively creating hypotheses after the data were analyzed) that are not disclosed fully. Researchers in organizational behavior and other fields have considerable latitude and "degrees of freedom" in terms of the choices, judgment calls, and trial-and-error decisions they make in every step of the research process—from theory, to design, measurement, analysis, and reporting of results (Bakker et al., 2012; Simmons et al., 2011). Consequently, other researchers are unable to reach similar conclusions due to insufficient information (i.e., low transparency) of what happened in what we label the "research kitchen" (e.g., Bakker et al., 2012; Bergh et al., in press; Bergh et al., 2017; Cortina, Green et al., 2017).

We are particularly interested in insufficient methodological transparency because it masks outright fraudulent acts (as committed by, for example, Hunton & Rose, 2011 and Stapel & Semin, 2007), serious errors (as committed by, for example, Min & Mitsuhashi, 2012 and Walumbwa, Luthans, Avey, & Oke, 2011), and questionable research practices (as described by Banks, Rogelberg, et al., 2016). Moreover, due to low methodological transparency, many of these errors are either never identified, or identified several years after publication. For example, it took at least four years to retract articles published in *Academy of Management Journal*, *Organization Science*, and other journals by now disgraced former University of Mannheim professor Ulrich Lichtenthaler. Greater methodological transparency could have substantially aided earlier discovery and possibly even prevented these and many other subsequently retracted manuscripts from being published by making clear that the data used were part of a larger dataset (e.g., Min & Mitsuhashi, 2012), providing information regarding decisions to include certain variables (e.g., Lichtenthaler, 2008), and being explicit about levels of inquiry and analysis (e.g., Walumbwa et al., 2011).

We pause here to make an important clarification. Our discussion of transparency, or lack thereof, does not mean that we wish to discourage discovery- and trial-and-error-oriented research. To the contrary, epistemological approaches other than the pervasive hypothetico-deductive model – which has become dominant in management and related fields since before World War II (Cortina, Aguinis, & DeShon, 2017) – are indeed useful and even necessary. For example, inductive and abductive approaches can lead to important theory advancements and discoveries (Bamberger & Ang, 2016; Fisher & Aguinis, 2017; Hollenbeck & Wright, 2016). Sharing our perspective, Hollenbeck and Wright (2016) defined "tharking" as "clearly and transparently presenting new hypotheses that were derived from post hoc results in the Discussion section of an article. The emphasis here is on how (transparently) and where (in the Discussion section) these actions took place" (p. 7). So, we are not advocating a rigid adherence to a hypothetico-deductive approach but, rather, epistemological and methodological plurality that has high methodological transparency.

References

Note: Sources preceded by an asterisk were used to generate recommendations in six tables included in our full-length manuscript, which include detailed recommendations on how to enhance methodological transparency (a) for researchers at the theory, design, measurement, analysis, and reporting stages, and (b) for reviewers, editors, and journals regarding the manuscript submission and review process

*Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology*, *90*, 94-107.

Aguinis, H., Cascio, W. F., & Ramani, R. S. (in press). Science's reproducibility and replicability crisis: International business is not immune. *Journal of International Business Studies*. doi:10.1057/s41267-017-0081-0

*Aguinis, H., Gottfredson R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16,* 270–301.

*Aguinis, H., & Vandenberg, R. J. (2014). An ounce of prevention is worth a pound of cure: Improving research quality before data collection. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 569-595.

*Aguinis, H., Werner, S., Abbott, J. L., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods*, *13*, 515-539.

*Alhija, F. N. A., & Levy, A. (2009). Effect size reporting practices in published articles. *Educational and Psychological Measurement*, *69*, 245-265.

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

*Aram, J. D., & Salipante, P. F. (2003). Bridging scholarship in management: Epistemological reflections. *British Journal of Management, 14*, 189-205.

*Aytug, Z. G., Rothstein, H. R., Zhou, W., & Kern, M. C. (2012). Revealed or concealed? Transparency of procedures, decisions, and judgment calls in meta-analyses. *Organizational Research Methods*, *15*, 103-133.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*, 543-554.

*Bakker, M., & Wicherts, J. M. (2011). The (mis) reporting of statistical results in psychology journals. *Behavior Research Methods*, *43*, 666-678.

*Balluerka, N., Gómez, J., & Hidalgo, D. (2005). The controversy over null hypothesis significance testing revisited. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*, 55-70.

Bamberger, P., & Ang, S. (2016). The quantitative discovery: What is it and how to get it published. *Academy of Management Discoveries, 2,* 1-6.

*Bampton, R, & Cowton, C. (2013). Taking stock of accounting ethics scholarship: A review of the journal literature. *Journal of Business Ethics, 114*, 549-563.

*Banks, G. C., O'Boyle, E. H., Pollack, J. M., White, C. D., Batchelor, J. H., Whelpley, C. E., Abston, K. A., Bennett, A. A., & Adkins, C. L. (2016). Questions about questionable research practices in the field of management. *Journal of Management, 42*, 5-20.

*Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016). Editorial:

Evidence on questionable research practices: The good, the bad, and the ugly. *Journal of Business and Psychology, 31*, 323-338.
*Bansal, P., & Corley, K. (2011). The coming of age for qualitative research: Embracing the diversity of qualitative methods. *Academy of Management Journal*, *54,* 233-237.
*Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, *61*, 1139-1160.
*Becker, T. E. (2005). Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. *Organizational Research Methods*, *8*, 274-289.
Bedeian, A. G., Taylor, S. G., & Miller, A. N. (2010). Management science on the credibility bubble: Cardinal sins and various misdemeanors. *Academy of Management Learning & Education*, *9*, 715-725.
*Bergh, D. D., Aguinis, H., Heavey, C., Ketchen, D. J., Boyd, B. K., Boyd, B. K., Su, P., Lau, C. L. L., & Joo, H. (2016). Using meta-analytic structural equation modeling to advance strategic management research: Guidelines and an empirical illustration via the strategic leadership-performance relationship. *Strategic Management Journal*, *37*, 477-497.
Bergh, D. D., Sharp, B. M., Aguinis, H., & Li, M. (in press). Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization.* Advance online publication. doi: https://doi.org/10.1177/1476127017701076
Bergh, D. D., Sharp, B., & Li, M. (2017). Tests for identifying" red flags" in empirical findings: Demonstration and recommendations for authors, reviewers and editors. *Academy of Management Learning & Education, 16,* 110-124.
*Bernerth, J. B., & Aguinis, H. (2016). A critical review and best-practice recommendations for control variable usage. *Personnel Psychology*, *69*, 229-283.
*Bettis, R. A. (2012). The search for asterisks: Compromised statistical tests and flawed theories. *Strategic Management Journal*, *33*, 108-113.
*Bettis, R. A., Ethiraj, S., Gambardella, A., Helfat, C., & Mitchell, W. (2016). Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal*, *37*, 257-261.
*Bluhm, D. J., Harman, W., Lee, T. W., & Mitchell, T. R. (2011). Qualitative research in management: A decade of progress. *Journal of Management Studies, 48*, 1866-1891.
*Bono, J. E., & McNamara, G. (2011). From the editors: Publishing in AMJ- Part 2: Research design. *Academy of Management Journal, 54,* 657-660.
*Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. (2016). HARKing's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psychology, 69*, 709-750.
*Boyd, B. K., Gove, S., & Hitt, M. A. (2005). Construct measurement in strategic management research: Illusion or reality?. *Strategic Management Journal*, *26*, 239-257.
*Brannick, M. T., Chan, D., Conway, J. M., Lance, C. E., & Spector, P. E. (2010). What is method variance and how can we cope with it? A panel discussion. *Organizational Research Methods*, *13*, 407-420.
*Brooks, M. E., Dalal, D. K., & Nolan, K. P. (2014). Are common language effect sizes easier to understand than traditional effect sizes?. *Journal of Applied Psychology*, *99*, 332-340.
*Brutus, S., Aguinis, H., & Wassmer, U. (2013). Self-reported limitations and future directions in scholarly reports: Analysis and recommendations. *Journal of Management, 39*, 48-75.
*Brutus, S., Gill, H., & Duniewicz, K. (2010). State of science in industrial and organizational

psychology: A review of self-reported limitations. *Personnel Psychology*, *63*, 907-936.
Byington, E. K. & Felps, W. (2017). Solutions to the credibility crisis in management science. *Academy of Management Learning & Education, 16,* 142-162.
*Cafri, G., Kromrey, J. D., & Brannick, M. T. (2010). A meta-meta-analysis: Empirical review of statistical power, type I error rates, effect sizes, and model selection of meta-analyses published in psychology. *Multivariate Behavioral Research*, *45*, 239-270.
*Capraro, R. M., & Capraro, M. M. (2002). Treatments of effect sizes and statistical significance tests in textbooks. *Educational and Psychological Measurement*, *62*, 771-782.
*Casper, W. J., Eby, L. T., Bordeaux, C., Lockwood, A., & Lambert, D. A. (2007). Review of research methods in IO/OB work--family research. *Journal of Applied Psychology, 92*, 28-43.
*Castro, S. L. (2002). Data analytic methods for the analysis of multilevel questions: A comparison of intraclass correlation coefficients, $r_{wg(j)}$, hierarchical linear modeling, within- and between-analysis, and random group resampling. *Leadership Quarterly*, *13*, 69-93.
*Chandler, G. N., & Lyon, D. W. (2001). Issues of research design and construct measurement in entrepreneurship research: The past decade. *Entrepreneurship: Theory and Practice, 25*, 101-114.
*Chang, S., van Witteloostuijn, A., & Eden, L. (2010). From the editors: Common method variance in international business research. *Journal of International Business Studies*, *41*, 178-184.
*Chenail, R. J. (2009). Communicating your qualitative research better. *Family Business Review*, *22*, 105-108.
*Combs, J. G. (2010). Big samples and small effects: Let's not trade relevance and rigor for power. *Academy of Management Journal, 53*, 9-13.
*Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods, 6,* 147-168.
*Conway, J. M., & Lance, C. E. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business & Psychology*, *25*, 325-334.
*Cools, E., Armstrong, S. J., & Verbrigghe, J. (2014). Methodological practices in cognitive style research: Insights and recommendations from the field of business and psychology. *European Journal of Work & Organizational Psychology, 23*, 627-641.
*Cortina, J. M. (2003). Apples and oranges (and pears, oh my!): The search for moderators in meta-analysis. *Organizational Research Methods*, *6*, 415-439.
Cortina, J. M., Aguinis, H., & DeShon, R. P. (2017). Twilight of dawn or of evening? A century of research methods in the Journal of Applied Psychology. *Journal of Applied Psychology, 102,* 274-290.
Cortina, J. M., Green, J. P., Keeler, K. R., & Vandenberg, R. J. (2017). Degrees of freedom in SEM: Are we testing the models that we claim to test? *Organizational Research Methods*, *20,* 350-378.
*Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement*, *61,* 229-248.
*Crook, T. R., Shook, C. L., Morris, M. L., & Madden, T. M. (2010). Are we there yet? An assessment of research design and construct measurement practices in entrepreneurship research. *Organizational Research Methods*, *13*, 192-206.

Davis, G. F. (2015). What is organizational research for? *Administrative Science Quarterly*, *60*, 179-188.
*Dionne, S. D., Gupta, A., Sotak, K. L., Shirreffs, K. A., Serban, A., Hao, C., Kim, D. H., & Yammarino, F. J. (2014). A 25-year perspective on levels of analysis in leadership research. *Leadership Quarterly, 25*, 6-35.
*Edwards, J. R. (2008). To prosper, organizational psychology should… Overcome methodological barriers to progress. *Journal of Organizational Behavior*, *29*, 469-491.
*Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*, 155-174.
*Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods*, *13*, 668-689.
*Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal, 50*, 25-32.
*Evert, R. E., Martin, J. A., McLeod, M. S., & Payne, G. T. (2016). Empirics in family business research. *Family Business Review, 29*, 17-43.
*Fan, X., & Thompson, B. (2001). Confidence intervals for effect sizes confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. *Educational and Psychological Measurement*, *61*, 517-531.
*Feng, G. C. (2015). Mistakes and how to avoid mistakes in using intercoder reliability indices. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 11*, 13-22.
*Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63*, 665-694.
*Finch, S., Cumming, G., & Thomason, N. (2001). Colloquium on effect sizes: The roles of editors, textbook authors, and the publication manual reporting of statistical inference in the Journal of Applied Psychology: Little evidence of reform. *Educational and Psychological Measurement*, *61*, 181-210.
*Firebaugh, G. (2007). Replication data sets and favored-hypothesis bias: Comment on Jeremy Freese (2007) and Gary King (2007). *Sociological Methods & Research*, *36*, 200-209.
Fisher, G., & Aguinis, H. (2017). Using theory elaboration to make theoretical advancements. *Organizational Research Methods, 20,* 438-464.
*Freese, J. (2007a). Overcoming objections to open-source social science. *Sociological Methods & Research*, *36*, 220-226.
*Freese, J. (2007b). Replication standards for quantitative social science: Why not sociology. *Sociological Methods & Research*, *36*, 153-172.
George, G. (2014). Rethinking management scholarship. *Academy of Management Journal*, *57*, 1-6.
*Gephart, R. P. (2004). Qualitative research and the Academy of Management Journal. *Academy of Management Journal, 47*, 454-462.
*Gerber, A. S., & Malhotra, N. (2008). Publication bias in empirical sociological research: Do arbitrary significance levels distort published results?. *Sociological Methods & Research*, *37*, 3-30.
*Geyskens, I., Krishnan, R., Steenkamp, J. M., & Cunha, P. V. (2009). A review and evaluation of meta-analysis practices in management research. *Journal of Management, 35*, 393-419.
*Gibbert, M., & Ruigrok, W. (2010). The 'what' and 'how' of case study rigor: Three strategies.

Organizational Research Methods, *13*, 710-737.
*Gibbert, M., Ruigrok, W., & Wicki, B. (2008). What passes as a rigorous case study?. *Strategic Management Journal, 29*, 1465-1474.
*Goldfarb, B., & King, A. A. (2016). Scientific apophenia in strategic management research: Significance tests & mistaken inference. *Strategic Management Journal*, *37*, 167-176.
Goodman, S. N., Fanelli, D., & Ioannidis, J. P. (2016). What does research reproducibility mean?. *Science Translational Medicine, 8*, 341-353.
*Haans, R. F., Pieters, C., & He, Z. L. (2016). Thinking about U: Theorizing and testing U-and inverted U-shaped relationships in strategy research. *Strategic Management Journal, 37*, 1177-1195.
*Hair, J. F., Sarstedt, M., Pieper, T. M., & Ringle, C. M. (2012). The use of partial least squares structural equation modeling in strategic management research: A review of past practices and recommendations for future applications. *Long Range Planning, 45*, 320-340.
*Hoekstra, R., Finch, S., Kiers, H. L., & Johnson, A. (2006). Probability as certainty: Dichotomous thinking and the misuse of *p* values. *Psychonomic Bulletin & Review*, *13*, 1033-1037.
*Hoetker, G. (2007). The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal*, *28*, 331-343.
*Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement, 60*, 523-531.
Hollenbeck, J. R. & Wright, P. M. (2016). Harking, sharking, and tharking: Making the case for post hoc analysis of scientific data. *Journal of Management, 43*, 5-18.
*Hubbard, R., & Ryan, P. A. (2000). Statistical significance with comments by editors of marketing journals: The historical growth of statistical significance testing in psychology—and its future prospects. *Educational and Psychological Measurement, 60,* 661-681.
Hunton, J. E., & Rose, J. M. (2011). Retraction. Effects of anonymous whistle-blowing and perceived reputation threats on investigations of whistle-blowing allegations by audit committee Members. *Journal of Management Studies, 48, 75-98.* (Retraction published 2015, *Journal of Management Studies, 52*, 849-849)
*Jackson, D. L., Gillaspy Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*, 6-23.
John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524-532.
Karabag, S. F., & Berggren, C. (2016). Misconduct, marginality and editorial practices in management, business and economics journals. *PloS one*, *11*, e0159492.
*Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measures designs: A review. *British Journal of Mathematical and Statistical Psychology*, *54*, 1-20.
*Klein, K. J., & Kozlowski, S. W. (2000). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods*, *3*, 211-236.
*Leavitt, K., Mitchell, T. R., & Peterson, J. (2010). Theory pruning: Strategies to reduce our dense theoretical landscape. *Organizational Research Methods*, *13*, 644-667.
*LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*, 815-852.
Lichtenthaler, U. (2008). Retraction. Externally commercializing technology assets: An

examination of different process stages. *Journal of Business Venturing, 23,* 445-464 (Retraction published 2013, *Journal of Business Venturing, 28*, 581).
*Mathieu, J. E, & Taylor, S. R. (2006). Clarifying conditions and decision points for mediational type inferences in organizational behavior. *Journal of Organizational Behavior*, *27*, 1031–1056.
*McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*, 64-82.
*McLeod, M., Payne, G., & Evert, R. (2016). Organizational ethics research: A systematic review of methods and analytical techniques. *Journal of Business Ethics, 134*, 429-443.
Min, J., & Mitsuhashi, H. (2012). Retraction. Dynamics of unclosed triangles in alliance networks: Disappearance of brokerage positions and performance consequences. *Journal of Management Studies 49,* 1078-1108 (Retraction published 2016, *Journal of Management Studies*, *51*, 868).
*Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, *5*, 241-301.
*Nuijten, M. B., Hartgerink, C. H., Assen, M. A., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods, 48,* 1205-1226.
O'Boyle, E. H., Banks, G. C., & Gonzalez-Mulé, E. (2017). The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management, 43,* 376-399.
*Petrocelli, J. V., Clarkson, J. J., Whitmire, M. B., & Moon, P. E. (2013). When ab≠ c–c′: Published errors in the reports of single-mediator models. *Behavior Research Methods*, *45*, 595-601.
*Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement, 64,* 916-924.
*Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management*, *36*, 94-120.
*Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879-903.
*Rogelberg, S., Adelman, M., & Askay, D. (2009). Crafting a successful manuscript: Lessons from 131 reviews. *Journal of Business & Psychology, 24,* 117-121
*Scandura, T. A., & Williams, E. A. (2000). Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal, 43*, 1248-1264.
*Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147-177.
*Schriesheim, C. A., Castro, S. L., Zhou, X. T., & Yammarino, F. J. (2002). The folly of theorizing "A" but testing "B": A selective level-of-analysis review of the field and a detailed leader–member exchange illustration. *Leadership Quarterly*, *12*, 515-551.
Schwab, A., & Starbuck, W. (2017). A call for openness in research reporting: How to turn covert practices into helpful tools. *Academy of Management Learning & Education, 16,* 124-141.
*Shen, W., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., & Ones, D. S. (2011). Samples in applied psychology: Over a decade of research in review. *Journal of Applied*

Psychology, 96, 1055-1064.
*Sijtsma, K. (2016). Playing with data: Or how to discourage questionable research practices and stimulate researchers to do things right. *Psychometrika, 81*, 1-15.
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.
*Smith-Crowe, K., Burke, M. J., Cohen, A., & Doveh, E. (2014). Statistical significance criteria for the r_{wg} and average deviation interrater agreement indices. *Journal of Applied Psychology, 99,* 239-261.
Stapel, D. A., & Semin, G. R. (2007). Retraction. The magic spell of language: Linguistic categories and their perceptual consequences. *Journal of Personality & Social Psychology, 93, 23-33. (*Retraction published 2013, *Journal of Personality & Social Psychology*, *104*, 197).
*van Aken, J. E. (2004). Management research based on the paradigm of the design sciences: The quest for field-tested and grounded technological rules. *Journal of Management Studies, 41,* 219-246.
*Van Iddekinge, C. H., & Ployhart, R. E. (2008). Developments in the criterion-related validation of selection procedures: A critical review and recommendations for practice. *Personnel Psychology, 61*, 871-925.
*Waldman, I. D., & Lilienfeld, S. O. (2016). Thinking about data, research methods, and statistical analyses: Commentary on Sijtsma's (2014) "Playing with Data". *Psychometrika, 81*, 16-26.
Walumbwa, F. O., Luthans, F., Avey, J. B., & Oke, A. (2011). Retraction. Authentically leading groups: The mediating role of collective psychological capital and trust. *Journal of Organizational Behavior, 32,* 4-24 (Retraction published 2014, *Journal of Organizational Behavior*, *35*, 746).
*Werner, S., Praxedes, M., & Kim, H. G. (2007). The reporting of nonresponse analyses in survey research. *Organizational Research Methods*, *10*, 287-295.
Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PloS one*, *6*, e26828.
*Wigboldus, D. H., & Dotsch, R. (2016). Encourage playing with data and discourage questionable reporting practices. *Psychometrika, 81,* 27-32.
*Wright, P. M. (2016). Ensuring research integrity: An editor's perspective. *Journal of Management*, *42*, 1037-1043.
*Yammarino, F. J., Dionne, S. D., Chun, J. U., & Dansereau, F. (2005). Leadership and levels of analysis: A state-of-the-science review. *Leadership Quarterly*, *16*, 879-919.
*Zhang, Y., & Shaw, J. D. (2012). From the editors: Publishing in AMJ-part 5: Crafting the methods and results. *Academy of Management Journal, 55,* 8-12.

Disentangling Reflection Strategy and Focus:

The Combined Effects of Task and Imaginative Reflection on Creativity

In recent years organizations have become increasingly focused on improving employees' creativity. Individual creativity has been depicted as a core component of personal and professional success (Goclowska & Crisp, 2014), and is a necessity for organizational growth, performance and long term survival (Anderson, Potocnik, & Zhou, 2014). Given the importance of individual creativity, the pursuit of strategies that may help develop these creative skills in employees has become a major domain of attention for practitioners (e.g., Levirne, 2015) and researchers (e.g., Reid, de Brentani, & Kleinschmidt, 2014). Drawing from their demonstrated success for improving task performance and their cost-effectiveness, creativity researchers have hinted at the important role reflection may play in the creative thinking process (e.g., Cohen & Ferrari, 2010; Gruber & Davis, 1988; Verhaeghen, Joormann, Aikman, & Shelley, 2014; Verhaeghen, Joormann, & Khan, 2005; Watkins, 2008). However, the holistic approach adopted in previous reflection intervention studies (see DeRue, Nahrgang, Hollenbeck, & Workman, 2012.) hinders the developments of straightforward predictions about the effects of reflection on creativity. Therefore, in the current paper, we propose adopting a building block approach to reflection interventions to disentangle the potential positive and negative effects of different reflection strategies on creativity.

On the basis of a theoretical analysis of reflection instructions used in previous research, we propose and test a new building block framework distinguishing two dimensions: reflection strategy (i.e., task reflection vs. imaginative reflection) and reflection focus (i.e., own performance vs. other's performance). Drawing on theory and insights in cognitive psychology on knowledge structures (e.g., Bilalic, McLead, & Gobet, 2008; Johnson-Laird, 1983; Rumelhart, 1984), we argue that the combined use of task reflection - a cognitive process through which individuals attempt to increase their awareness of previous experiences, by analyzing the experiences and concomitant behavior, and therefore their ability to learn from them (Gordon & Smith Hullfish, 1961) - and imaginative reflection – the cognitive process whereby people are challenged to think about "What might have happened if another approach was chosen?" (Ellis, Carette, Anseel, & Lievens, 2014) - may offset potential drawbacks each of them might have in isolation. To better understand the differential effects each strategy might have, we advance reflection focus as a second dimension in our framework (for example see Ellis, Ganzach, Castle, & Sekely, 2010). We theorize that each of the two reflection strategies may risk undermining creativity when applied to a particular reflection focus. More precisely, in line with the cognitive entrenchment perspective of Dane (2010), we expect that task reflection will be less effective when reflecting on one's own performance whereas imaginative reflection will be less effective when reflecting on other's performance. However, by combining the

two strategies, the potential drawbacks of one reflection strategy will be compensated by the strengths of the other one.

To test our ideas, we report three experiments to disentangle the influence of reflection strategy and focus on creativity. In the first study, an experiment was conducted with participants of different healthcare organizations. Two hundred and fifty five employees were randomly assigned to a 2 (task reflection versus no task reflection; between persons) x 2 (imaginative reflection versus no imaginative reflection; between persons) x 2 (creative task performance T1 versus creative task performance T2; within persons) mixed design, yielding four between-person conditions: task reflection, imaginative reflection, combined reflection, and a control condition. In study two, 191 participants (psychology students) were randomly assigned to a 2 (task reflection versus no reflection; between persons) x 2 (imaginative reflection versus no imaginative reflection; between persons) design with creative performance as dependent variable. Experiment 1 focused on reflection on one's own performance whereas experiment 2 focused on reflection on other's performance. In the third experiment, 255 participants (psychology students) were randomly assigned to a 2 (own versus other's performance; between persons) x 2 (task reflection versus imaginative reflection; between persons) design, testing our fully crossed preliminary 2 by 2 framework on creative performance.

Results generally supported our idea that a combination of reflection strategies is most effective to support creative performance. Interaction effects between reflection strategy and reflection focus provided partial support for the assumed underlying cognitive mechanisms. Both Study 1 and Study 3 showed that, when reflecting on their own performance, participants were more creative when they engaged in imaginative reflection instead of task reflection. Surprisingly, we did not find a significant difference between combined reflection conditions and control conditions in study 1 and 2, suggesting that reflection overall did not improve creativity. However, an inappropriate reflection intervention can hinder creativity improvement on the workplace. For example, Study 1 showed task reflection on one's own performance had a detrimental effect on creativity. We discuss the threat of a "pro-reflection" bias in the literature and suggest that not all reflection interventions will be uniformly useful, especially for creativity. We delineate future research avenues to further expand on the proposed building block framework.

Anderson, N., Potocnik, K., & Zhou, J. 2014. Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. ***Journal of Management,*** 40: 1297–1333.

Bilalic, M., McLeod, P., & Gobet, F. 2008. Inflexibility of experts—Reality or myth? Quantifying the Einstellung effect in chess masters. ***Cognitive Psychology,*** 56: 73–102.

Cohen, J. R., & Ferrari, J. R. 2010. Take some time to think this over: The relation between rumination, indecision, and creativity. ***Creativity Research Journal,*** 22: 68–73.

Dane, E., Baer, M., Pratt, M. G., & Oldham, G. R. 2011. Rational versus intuitive problem solving: How thinking "off the beaten path" can stimulate creativity. ***Psychology of Aesthetics, Creativity, and the Arts,*** 5: 3–12.

DeRue, D. S., Nahrgang, J. D., Hollenbeck, J. R., & Workman, K. 2012. A quasi-experimental study of after-event reviews and leadership development. ***Journal of Applied Psychology,*** 97: 997–1015.

Ellis, S., Carette, B., Anseel, F., & Lievens, F. 2014. Systematic reflection: Implications for learning from failures and successes. ***Current Directions in Psychological Science,*** 23: 67–72.

Ellis, S., Ganzach, Y., Castle, E., & Sekely, G. 2010. The effect of filmed versus personal after-event reviews on task performance: The mediating and moderating role of self-efficacy. ***Journal of Applied Psychology,*** 95(1): 122.

Goclowska, M. A., & Crisp, R. J. 2014. How dual-identity processes foster creativity. ***Review of General Psychology,*** 18: 216–236.

Gordon, H., & Smith Hullfish, P. H. 1961. ***Reflective thinking: The method of education.*** New York: Dodd, Mead & Co.

Gruber, H. E., & Davis, S. N. 1988. Inching our way up Mount Olympus: The evolving-system approach to creative thinking. In R. J. Sternberg (ed.), ***The Nature of Creativity:*** 243–270. New York: Cambridge University Press

Johnson-Laird, P. N. 1983. ***Mental models: Towards a cognitive science of language, inference, and consciousness.*** Cambridge, MA: Harvard University Press.

Levirne, J. 2005 (May 15). ***The most overlooked way of stimulating team creativity.*** Retrieved from https://hbr.org/2015/05/the-most-overlooked-way-of-stimulating-team-creativity

Reid, S. E., de Brentani, U., & Kleinschmidt, E. J. 2014. Divergent thinking and market visioning competence: An early front-end radical innovation success typology. ***Industrial Marketing Management,*** 43: 1351–1361.

Rumelhart, D. E. 1984. Schemata and the cognitive system. In R. S. Wyer & T. K. Srull (Eds.), ***Handbook of social cognition*****:** 161–188. Hillsdale, NJ: Lawrence Erlbaum Associates.

Verhaeghen, P., Joormann, J., Aikman, S. N. 2014. Creativity, mood, and the examined life: Self-reflective rumination boosts creativity, brooding breeds dysphoria. ***Psychology of Aesthetics Creativity and the Arts,*** 8: 211–218.

Verhaeghen, P., Joormann, J., & Khan, R. 2005. Why we sing the blues: The relation between self-reflective rumination, mood, and creativity. ***Emotion,*** 5: 226–232.

Watkins, E. R. 2008. Constructive and unconstructive repetitive thought. ***Psychological Bulletin,*** 134: 163-206.

A new level of analysis in climate research: Informal friends cliques

1. Literature Review

It is well accepted that organizational phenomena are inherently multilevel as opposed to occurring at a single level (Chan, 1998). Therefore, assessments of inter rater agreement, which are a precondition to aggregate data into higher levels (Chan, 1998; James, 1982; James, Choi, Ko, McNeil, Minton, Wright, & Kim, 2008), have become increasingly important over the last decade (Smith-Crowe, Burke, Cohen & Doveh, 2014). *Inter rater agreement* (i.e., agreement) reflects the extent to which raters provide the same rating (Cohen, Doveh, & Nahum - Shani, 2007) for a selected dimension, and is indicative of a collective interpretive process (Kozlowski & Hattrup, 1992) . However, many researchers have reported the existence of groups characterized by absence of agreement within their databases (for example: McKay, Avery, & Morris, 2008; Zohar, 2002; Zohar & Luria, 2004). These researchers had to contend with the decision to include or exclude these groups in the sample. Each decision has different consequences that may bias results (LeBreton & Senter, 2007). This is particularly valid for climate research, in which the acceptable level of analysis is the formal organizational unit (Ostroff, Kinicki, & Muhammad, 2013), and absence of agreement is often found. While scholars have argued that absence of agreement may be caused by the aggregation of multiple **smaller cliques (subgroups), which have a substantial agreement** within them and high variability between them (Moritz & Watson, 1998. p. 296). This claim has, to date, remained unexamined.

Friends clique may create share safety climate perceptions (an agreement), as a result of three social process: *Cognitive balance* (Heider, 1946), which causes friends to hold similar attitudes to their friend in order to reduce tension; *Assimilation,* which causes individuals to adopt friends' perceptions (e.g., Schulte, Cohen, & Klein, 2012); *Homophily*, which causes individuals to select their friends based on similarity (McPherson, Smith-Lovin, & Cook, 2001). In other words: individuals form ties with similar others and reach out to teammates who hold similar perceptions (Schulte et al., 2012). They then "catch" their friends' and their friends' friends attitudes. All these processes create the infrastructure for the argument that cliques may characterized by high level of agreement, and that different cliques may hold different perceptions of climate, all of which lead to reduced overall agreement within a formal unit.

2. Research's aim and hypotheses

The current research aims to introduce the informal friendship clique as a new level of analysis in climate research. This level of analysis may explain absence of agreement in formal units. We hypothesized that cliques will be characterized by high agreement/similar perceptions of safety climate; and within formal teams that suffer from absence of agreement, high level agreement cliques - will be found.

3. Methodology

The sample included 567 employees nested within 46 teams in industrial organizations. Network method was used to identify cliques. RWG_j (James, Damaree, & Wolf, 1984); AD_j (Burke, Finkelstein & Dusig, 1999); and ICC(1) (Bartko, 1976; James,

1982); and uniformity (González- Romá & Hernández, 2014) measurements, were used to assess level of agreement in safety climate within formal teams and within cliques.

4. Results

Sufficient levels of safety climate agreement indicated an existence of a clique level of analysis (RWG_j>.7: mean=.74; median=.9); AD_j<.8: mean= .79; median=.8); and ICC(1) were significant and higher than .05: ICC(1)=.07; $F(1,547)=1.372$, $p<.05$). 64.41 percents of the cliques presented a uniform pattern of perceptions. Moreover, 75 percent and above, of the formal teams characterized by absence of agreement - demonstrated high levels of agreement within at least one of their cliques.

5. Implications

Our findings demonstrate the importance of informal relation identification within formal teams. Friends' cliques level of analysis may be used to understand teams processes and creation of attitudes. Moreover it provides additional information that can foster decision whether to include teams within research sample or to drop them.

6. References

Bartko, J. J. (1976). On various intra class correlation reliability coefficients. *Psychological Bulletin, 83*(5), 762-765.

Burke, M. J., Finkelstein, L. M., & Dusig, M. S. (1999). On average deviation indices for estimating interrater agreement. *Organizational Research Methods, 2*(1), 49-68.

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83*(2), 234-246.

Cohen, A., Doveh, E., & Nahum-Shani, I. (2009). Testing agreement for multi-item scales with the indices rWG (J) and AD m (J). *Organizational Research Methods, 12*(1), 14s8-164.

Gonzlez-Rom, V., & Hernndez, A. (2014). Climate uniformity: Its influence on team communication quality, task conflict, and team performance. *Journal of Applied Psychology, 99*(6), 1042-1058.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology, 21*(1), 107-112.

James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology, 67*(2), 219-229.

James, L. R., Choi, C. C., Ko, C. E., McNeil, P. K., Minton, M. K., Wright, M. A., & Kim, K. (2008). Organizational and psychological climate: A review of theory and research. *European Journal of Work and Organizational Psychology,17*(1), 5-32.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69*(1), 85-98.

Kozlowski, S. W., & Hattrup, K. (1992). A disagreement about within-group agreement: Disentangling issues of consistency versus consensus. *Journal of Applied Psychology, 77*(2), 161-167.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*(4), 815-852.

McKay, P. F., Avery, D. R., & Morris, M. A. (2008). Mean racial□ethnic differences in employee sales performance: The moderating role of diversity climate. *Personnel Psychology, 61*(2), 349-374.

McPherson, M., Smith- Lovin, L., & Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(1), 415-444.

Moritz, S. E., & Watson, C. B. (1998). Levels of analysis issues in group psychology: Using efficacy as an example of a multilevel model. *Group Dynamics: Theory, Research, and Practice, 2*(4), 285-298.

Ostroff, C., Kinicki, A. J., & Tamkins, M. M. (2013). Organizational culture and climate. In I. B. Weiner (Ed.), *Handbook of psychology* (Second Edition ed., pp. 643-676). Hoboken, New Jersey: John Wiley & Sons, Inc.

Schulte, M., Cohen, N. A., & Klein, K. J. (2012). The coevolution of network ties and perceptions of team psychological safety. *Organization Science, 23*(2), 564-581.

Smith-Crowe, K., Burke, M. J., Cohen, A., & Doveh, E. (2014). Statistical significance criteria for the r WG and average deviation interrater agreement indices. *Journal of Applied Psychology, 99*(2), 239-261.

Zohar, D. (2002). The effects of leadership dimensions, safety climate, and assigned priorities on minor injuries in work groups. *Journal of Organizational Behavior, 23*(1), 75-92.

Zohar, D., & Luria, G. (2004). Climate as a social-cognitive construction of supervisory safety practices: Scripts as proxy of behavior patterns. *Journal of Applied Psychology, 89*(2), 322-333.

A Failure to Disagree:

The Role of Panel Diversity in Creative Evaluation and Forecasting

New ideas often spark controversy. Pioneers of personal computers were met by both frantic enthusiasm and discouraging dismissals, before gaining popularity. Monet's painting *Impression, soleil levant* was initially perceived as scandalous, and only later acknowledged for its seminal role in the development of the Impressionist movement. Even without inconveniencing landmarks in the history of science and art, every scholar can recollect having witnessed how the same idea can trigger inconsistent evaluations over its potential value and originality. Empirical studies on peer-reviews report strikingly high levels of disagreement among reviewers of scientific manuscripts (range ICC: 0.05 - 0.54; Miller, 2006; Starbuck, 2003, 2016), and grant applications (range r: 0.16 - 0.48; Marsh, Jayasinghe, & Bond, 2008).

Panels of independent experts are however expected to be immune to such controversies when ascertaining the creativity of an idea (e.g. Kaufman, Baer, Cole, & Sexton, 2008; Kaufman & Baer, 2012). The long-standing assumption behind this notion is that ideas possess a certain degree of novelty and usefulness, which can be observed and consensually measured by a group of appropriate judges (Amabile, 1982). Consensus is to be expected because: "to the extent that the task presented to subjects draws on special talents or experience-related skills […] the assessment will be insensitive to social-psychological effects" (1982, p. 1000).

As a consequence, it is common practice to employ panels of independent judges, who are expert in the domain of interest, to identify the best ideas (e.g. Girotra, Terwiesch, & Ulrich, 2010) and to predict their success (e.g. Berg, 2016; Kornish & Ulrich, 2014). A high degree of interrater agreement and reliability in their evaluations is usually used as measure of construct validity and thus posed as necessary requirement (Hennessey, Amabile, & Mueller, 2011).

We set to revisit these commonly held beliefs, in light of recent findings on creative evaluation and forecasting (e.g. Berg, 2016; Loewenstein & Mueller, 2016; Zhou, Wang, Song, & Wu, 2016). Specifically, we suggest that for creativity evaluations: *i)* domain expertise does not shield a panel from socio-psychological influences; *ii)* consensus might be a signal of lack of diversity, rather than domain expertise; and *iii)* panel expertise diversity is a better predictor of predictive accuracy than domain expertise.

We explore this alternative perspective on antecedents of panel creativity evaluations and forecasting accuracy by first investigating whether *domain expertise* indeed prevent *cultural factors* and *personality traits* from interfering with individual evaluation of creativity. We do this by attempting to replicate effects on creativity judgments of regulatory focus (Zhou et al., 2016), construal level (Mueller et al., 2014) openness to experience (Sijbom et al., 2015) and nationality (Loewenstein & Mueller, 2016), and systematically testing for a moderating role of domain expertise. Expanding this line of thought to the panel level, we explore whether *personality, cultural* and *expertise diversity* predicts the degree of consensus in panels' creativity judgments. We ground this hypothesis on previous work on deep-level diversity and creativity, which highlighted how diverse knowledge and perspectives might prevent individuals from reaching consensus (e.g. Harvey, 2013; Hoever, van Knippenberg, van Ginkel, & Barkema, 2012).

Finally, we explore whether *panel expertise diversity* is positively related to the predictive validity of its creativity judgments. Since panels with diverse expertise are more favourable towards novel ideas (Criscuolo, Dahlander, Grohsjean, & Salter, 2016), and domain expertise does not seem to outperform general knowledge in creative forecasting (cf. Berg, 2016; Dailey & Mumford, 2006; Kornish & Ulrich, 2014), we suggest that panels composed of employees from different sectors may be more accurate in identifying the most successful ideas.

We tested our hypotheses in a first exploratory study, recruiting employees from multiple industries and countries (n = 128, mean working years = 10.14, mean age =32.77, proportion women = 0.45) via an online platform. Participants evaluated the creativity and forecasted the commercial potential of a set of health-care product concepts recently launched on Kickstarter. To test our hypotheses at the panel level, we randomly sampled respondents to simulate one thousand panels of independent judges (cf. Kornish & Ulrich, 2014), and aggregated individual responses to construct panel attributes and evaluations. Despite its limitations, this method

allows to economically obtain necessary information on a substantial number of panels and to disentangle within-idea and between-judges variance (an endeavour otherwise unfeasible through meta-analytical efforts or archival data).

Preliminary results seem to support our hypotheses. Although we fail to replicate the effects of openness to experience and construal level, promotion focus and nationality are found to significantly affect individuals' mean creativity judgments. Importantly, the systematic bias is equally strong for domain experts than for employees with lower or different expertise. We also observe from the partitioning of variance that between-judges effects explain variance in creativity judgments above and beyond between-ideas effects. This evidence not only supports the claim that biases against creativity may indeed be generalizable to experts, but also it raises the question of whether we should still conceptualize creativity as a characteristic of an idea, rather than as a perception of those bound to judge it.

Also consistently with hypotheses, panel diversity in regulatory foci, nationality, and expertise were observed to be negatively related to panel agreement in creativity judgments. Furthermore, panel expertise diversity, rather than panel consensus, appears to be positively related to a panel's accuracy in creative forecasting, but the former relation is contingent on the decision rule applied to aggregate individual evaluations (Csaszar & Eggers, 2013). Although more empirical evidence is needed to understand the link between diversity, consensus and forecasting, the study already shows that high degree of interrater agreement is indicative of more homogeneous perspectives, but not necessarily of more accurate evaluations. It also raises attention of fellow creativity scholars over the value of treating within-panel agreement as a variable of interest, rather than as statistical prerequisite. Finally, the findings have important practical implications, by highlighting how panel composition and decision rules may affect creativity evaluation and forecasting.

References

Amabile, T. M. (1982). Social Psychology of Creativity: A Consensual Assessment Technique. *Journal of Personality and Social Psychology*, *43*(5), 997–1013. http://doi.org/10.1037/0022-3514.43.5.997

Berg, J. M. (2016). Balancing on the Creative Highwire: Forecasting the Success of Novel Ideas in Organizations. *Administrative Science Quarterly*, *61*(3), 433–468. http://doi.org/10.1177/0001839216642211

Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. (2016). Evaluating Novelty: The Role of Panels in the Selection of R&D Projects. *Academy of Management Journal*. http://doi.org/10.5465/amj.2014.0861

Csaszar, F. A., & Eggers, J. P. (2013). Organizational Decision Making : An Information Aggregation View. *Management Science*, *59*(10), 2257–2277. http://doi.org//10.1287/mnsc.1120.1698

Dailey, L., & Mumford, M. D. (2006). Evaluative Aspects of Creative Thought: Errors in Appraising the Implications of New Ideas. *Creativity Research Journal*, *18*(3), 367–384. http://doi.org/10.1207/s15326934crj1803_11

Girotra, K., Terwiesch, C., & Ulrich, K. T. (2010). Idea Generation and the Quality of the Best Idea. *Management Science*, *56*(4), 591–605. http://doi.org/10.1287/mnsc.1090.1144

Harvey, S. (2013). A different perspective: The multiple effects of deep level diversity on group creativity. *Journal of Experimental Social Psychology*, *49*(5), 822–832. http://doi.org/10.1016/j.jesp.2013.04.004

Hennessey, B., Amabile, T. M., & Mueller, J. S. (2011). *Consensual Assessment*. *Encyclopedia of Creativity (Second Edition)* (2nd ed., Vol. 1). Elsevier Inc. http://doi.org/http://dx.doi.org/10.1016/B978-0-12-375038-9.00046-7

Hoever, I. J., van Knippenberg, D., van Ginkel, W. P., & Barkema, H. G. (2012). Fostering Team Creativity: Perspective Taking as Key to Unlocking Diversity's Potential. *Journal of Applied Psychology*, *97*(5), 982–996. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22774764

Kaufman, J. C., & Baer, J. (2012). Beyond New and Appropriate: Who Decides What Is Creative? *Creativity Research Journal*, *24*(1), 83–91. http://doi.org/10.1080/10400419.2012.649237

Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A Comparison of Expert and Nonexpert Raters Using the Consensual Assessment Technique. *Creativity Research Journal*, *20*(2), 171–178. http://doi.org/10.1080/10400410802059929

Kornish, L. J., & Ulrich, K. T. (2014). The Importance of the Raw Idea in Innovation: Testing the Sow's Ear Hypothesis. *Journal of Marketing Research (JMR)*, *51*(1), 14–26. http://doi.org/10.1509/jmr.12.0401

Loewenstein, J., & Mueller, J. S. (2016). Implicit Theories of Creative Ideas: How Culture Guides Creativity Assessments. *Academy of Management Discoveries*.

Mueller, J. S., Wakslak, C. J., & Krishnan, V. (2014). Construing creativity: The how and why of recognizing creative ideas. *Journal of Experimental Social Psychology*, *51*, 81–87. http://doi.org/10.1016/j.jesp.2013.11.007

Sijbom, R. B. L., Janssen, O., & Van Yperen, N. W. (2015). How to get radical creative ideas into a leader's mind? Leader's achievement goals and subordinates' voice of creative ideas. *European Journal of Work and Organizational Psychology*, *24*(2), 279–296. http://doi.org/10.1080/1359432X.2014.892480

Zhou, J., Wang, X. M., Song, L. J., & Wu, J. (2016). Is it New? Personal and Contextual Influences on Perceptions of Novelty and Creativity. *Journal of Applied Psychology*. http://doi.org/10.1037/apl0000166

The EDEN model: A framework for studying the dynamics of creative exploration at high resolution

Introduction

I present here two paradigms and a unified model for studying creative exploration. My goal is to describe a general approach for studying creative exploration at the *nano* level – analyzing and modeling creative behavior at a higher resolution than the typical micro level of analysis in OB. The presented paradigms apply a physics-like approach: studying simple model systems that allow for exact measurement and mathematical modeling of creative behavior. The first model system analyzes the motion of a pair of actors improvising together. The second model analyzes the behavior of individuals exploring a bounded visual space. The Explore-Discover-Enjoy-Next (EDEN) model sets the results from the two model system within a unified framework.

The presented approach was developed over the last six years at the Theater Lab at the Weizmann Institute of Science. The Theater Lab is an interdisciplinary research group employing tools from physics, computer science and psychology to study performance related phenomena such improvisation and creativity.

The Mirror Game: a quantitative paradigm for studying joint improvisation

Joint improvisation, where a group of people creates together without a pre-defined script, is usually associated with performance genres such as Jazz. However, other activities such as emergency management (Mendonca, Beroggi, & Wallace, 2001) or medical operations (Kneebone, 2011) can be considered as instances of joint improvisation. In order to study the basic mechanisms of joint improvisation, we employed the mirror game, a common practice in theater and dance (Noy, 2014). In the mirror game practice, two people imitate each other, creating synchronized and dance-like motion, with or without a pre-defined leader. We developed an experimental paradigm of the mirror game by reducing the motion to a single dimension (Noy, Dekel, & Alon, 2011). Participants imitate each other by moving handles on two linear tracks (Fig 1A, 1B), with the goal of creating together motion that is both synchronized and interesting. Expert improvisers are able to create complex and highly synchronized motion together (Fig 1C).

The mirror game (MG) experimental paradigm allowed us to pin down basic mechanism of joint improvisation: the ability of expert pairs to share leadership by mutually predicting each other's actions. Expert pairs can enter periods of synchronized motion, without the typical jitter that marks a reactive leader-follower relationship seen in novices (Hart, Noy, Feniger-Schaal, Mayo, & Alon, 2014; Noy et al., 2011) (see Fig 1D). This interpretation is in line with previous literature analyzing single person manual tracking (Weir, Stein, & Miall, 1989) and is corroborated by modeling two reactor-predicator controllers in a mirror configuration (Dahan, Noy, Hart, Mayo, & Alon, 2016; Noy et al., 2011). These co-predictive, jitter-less and synchronized periods were termed co-confident (CC) motions, and are suggested as a kinematic marker for the experience of *togetherness* in the MG (Noy, Levit-Binun, & Golland, 2015). Expert improvisers are able to enter and exit (Dahan et al., 2016) these moments of togetherness during joint improvisation.

The Creative Foraging Game: a quantitative paradigm for studying creative exploration

Creative exploration and discovery processes are central in science, art and technological innovation (Drucker, 2014; Kozbelt, 2006; Simonton, 1999). The creative foraging game was developed to address a need for quantitative paradigms for measuring creative exploration at high resolution. Current paradigms employ search spaces that are not enumerated and lack a clear metric. For example, in a common creativity test participants are asked to name alternate usages for a common object (e.g., a brick), and their response is manually coded for fluency and category switches. However, it is not clear how many possible solutions exist; what are the possible solutions categories; and what is the topology of the explored space – for example, what are the shortest paths between two solutions. In contrast, we study creative exploration in a bounded and well-defined visual space, building on previous works on studying human foraging at high resolution (Hills, Jones, & Todd, 2012; Hills, Todd, & Goldstone, 2015).

In the creative foraging game (CFG) people seek interesting and beautiful shapes in a bounded and well-defined geometric space (Fig 2A). Participants can move ten adjacent squares on a computer screen to create different shapes. Participants move one square at a time to a new position, and add shapes they find interesting and beautiful into a gallery. The timing of all moves and gallery choices is recorded. The setup therefore captures all intermediate steps as people explore a large (~36K) but finite metric space of possibilities (Noy et al., 2012).

Participants alternate between phases of exploration and exploitation: in exploration phases they meander, moving from shape to shape along non-minimal paths; in exploitation they move directly along minimal paths to collect shapes that are related within a meaningful theme. After a while, they return to exploration and discover new themes (Hart et al., 2017) (Fig 2B). Different people discover the same themes (Fig 2C), but enter them through transition shapes that are unique to each person (Fig 2D, 2E). These transition shapes are akin to mini creative leaps, in which people seem to suddenly comprehend the potential for a new type of shapes, grasping, in the words of Peter F. Drucker, ”what’s visible—but not yet seen” (Edersheim & Drucker, 2007).

The Explore-Discover-Enjoy-Next (EDEN) Model

The aforementioned results suggest a similar process in the two model systems: individuals and dyads move from exploration to exploitation phases via points of discovery (MG: switch from leader-follower dynamics to co-confident motion; GFC: entering a cluster of similar shapes through a non-prototypical shape). They then switch back from exploitation to exploration, deciding at some point to ‘move on’ (MG: exit from synchrony; CFG: leaving a cluster of similar shapes). In both paradigms exploitation phases are top-down and goal directed (MG: feed-forward co-predictive control; CFG: honing on a specific meaning field), while exploration phases are more reactive, with less forward planning.

A four-stage model, the Explore-Discover-Enjoy-Next (EDEN) model, captures this common process (Figure 3). The model grounds a common metaphor of creative search: exploration as walking in a ‘fog’, punctuated by discovery moments that bring clarity and forward planning. These moments of clarity might be related for the positive experience of flow and group flow. For example, during periods of co-predictive motion in the MG, participants report a higher sense of togetherness and show a significant increase in the heart rates

(Noy, Levit-Binun, et al., 2015). We therefore use the work Enjoy to describe this stage rather than the more common Exploit.

A theoretical contribution of the EDEN model is emphasizing the importance of the Next (move-on) stag - the point in which a creative explorer (or a team of) decides to leave a stable region of the search space, in order to start exploring again. For example, in the field of synchronized motion there is ample research on the conditions that brings people to *enter* synchrony (Repp & Su, 2013). However, we were the first to measure and mathematically model the phenomena of *exit* from synchrony (Dahan et al., 2016). A similar phenomenon might occur in the CFG: players leave a category of solutions (e.g. 'airplanes') way ahead of depletion, using a dynamics that cannot be explained according to the theory of Optimal Foraging (Hart et al., 2017). This suggests that the decision to exit an Enjoy/Exploit phase is not guided by the lack of resources. Future studies could clarify the conditions that bring explorers to advance from stage to stage in the EDEN model. These conditions might also be the results of individual exploration tendencies, as briefly discussed in the next section.

Relevance for OB studies

I present here an example of a possible connection of the aforementioned works and OB studies.

A reoccurring theme of the presented works are individual differences in exploration (Hart et al., 2014; Noy, Alon, & Friedman, 2015). An intriguing result from the CFG is a correlation between the average duration of the exploration phase and the average duration of the exploitation phase (Hart et al., 2017). People seem to span an axis between those who prefer short exploration and short exploitation, and those who prefer long exploration and long exploitation.

These individual tendencies might be present in creative processes in organizations, for example in brain-storming sessions. A testable hypothesis is that people who exhibit the short/short pattern in the CFG are quicker to suggest ideas in brain-storming sessions; while people who exhibit the long/long pattern in the CFG elaborate more on their ideas. If so, a possible effective method of allocating creative teams might be to mix people from these different types of creative exploration 'personalities'. Furthermore, if the observed axis of short/short vs long/long indeed reflects a stable personal tendency, it would be interesting to try to correlate these exploration phenotypes with personality traits such as the Big-5 traits (e.g., openness to experience) and attachment styles. Related to this idea, we have recently shown the people with a secure attachment style behave in a more exploratory way in the mirror game paradigm (Feniger-Schaal et al., 2016).

I hope to present our high-resolution approach for studying creative exploration in the upcoming IOBC conference, and explore further connections to the field of OB with the participating experts.

Figures

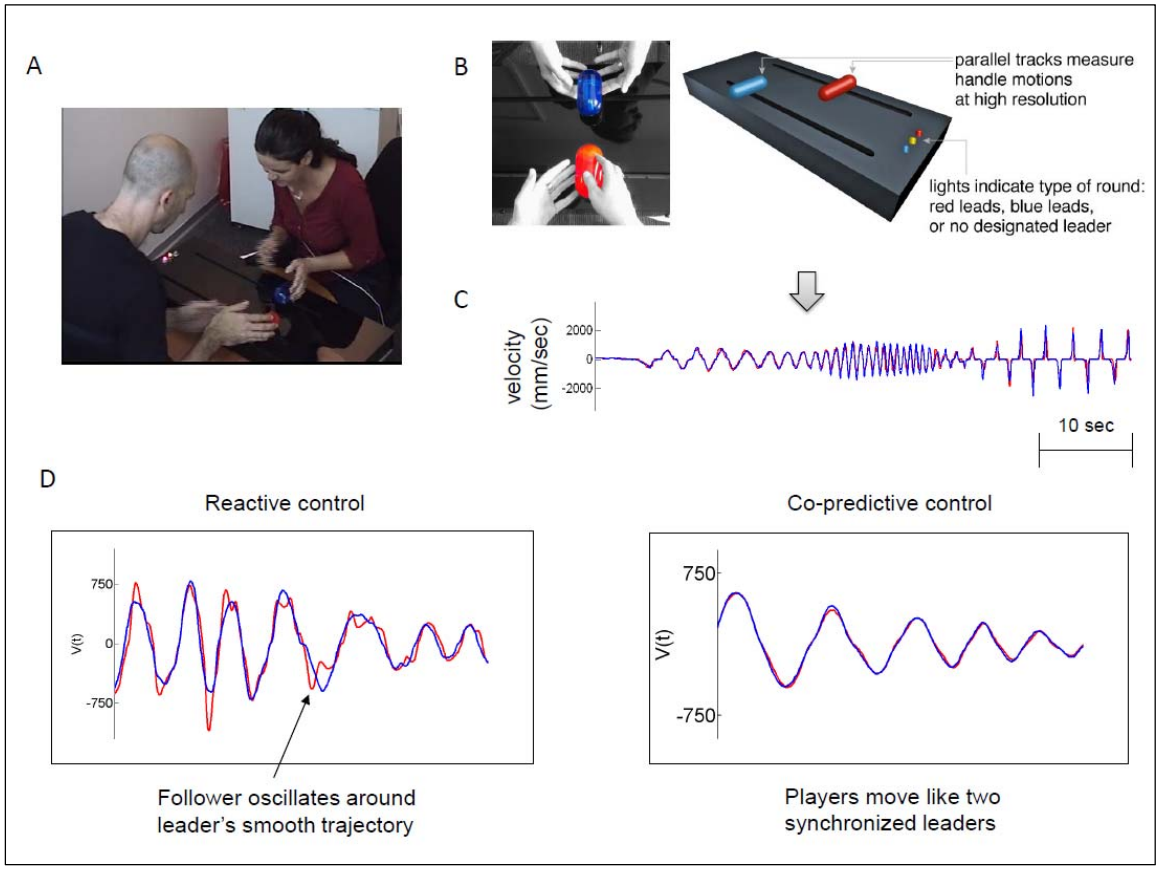

Figure 1. **The mirror game: a paradigm for studying joint improvisation.** A) Two players play a simplified version of the MG. Players imitate each other, with or without a designated leader. Players are instructed to "create synchronized and interesting motion together". B) The one-dimensional MG device. Players move handles along a 60 cm track while their motion is being tracked. C) An example of players' velocity traces. Shown is a 1 min round (no designated leader condition). The blue and red lines show velocity traces of the corresponding player. The produced motion goes through both gradual (eg increased frequency) and abrupt changes. D) Kinematic signatures of reactive (leader/follower) and co-predictive motion. Left: a leader (blue) produces smooth ('confident') motion, while the follower (red) follows the general trace of the leader, with an additional 1-2 Hz jittery motion. This type of jittery motion is considered as a marker of reactive control. Right: expert improvisers can enter a state in which they both produce synchronized and jitter-less motion, suggesting that they apply co-predictive control ('co-confident' motion). These periods are suggested to serve as a kinematic marker for the state of togetherness in the MG.

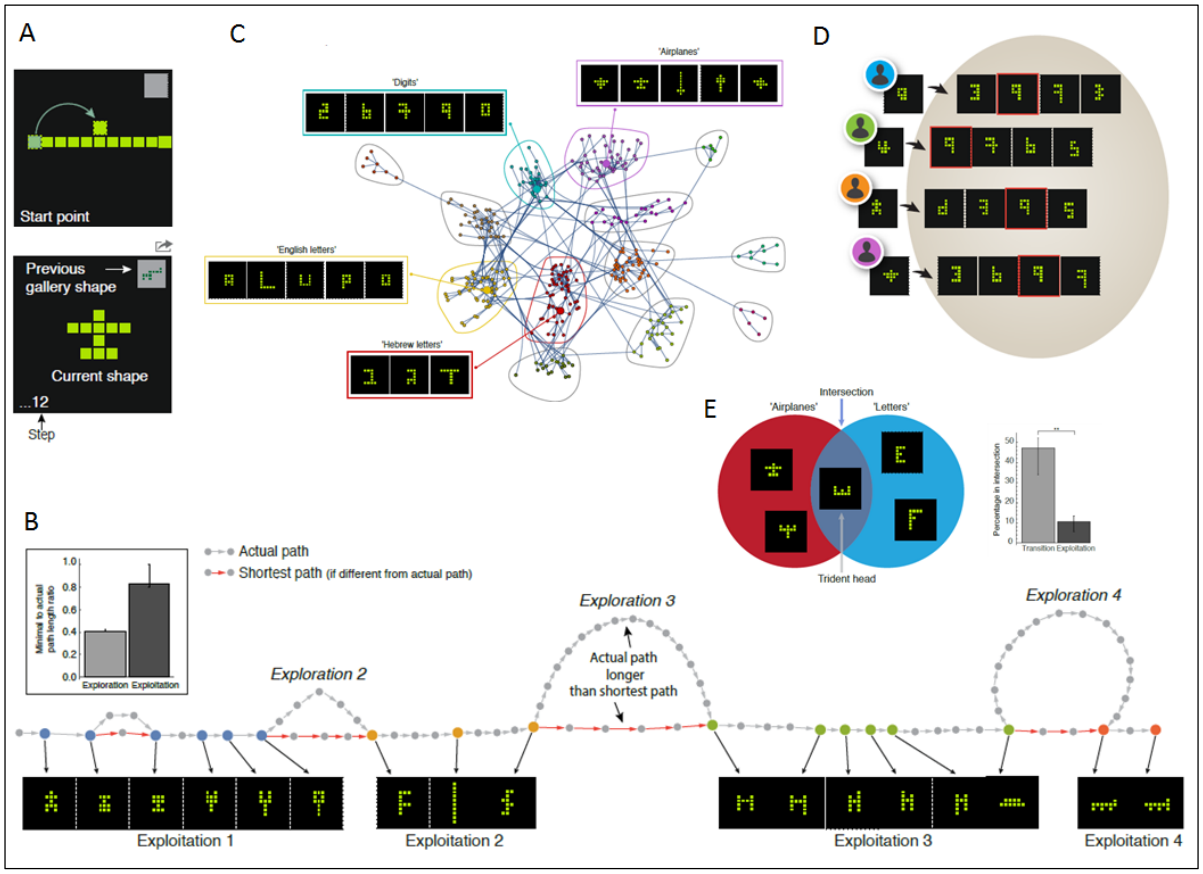

Figure 2. The creative foraging: a paradigm for studying creative exploration. A) Participants play a computer game in which they move squares and seek shapes which are 'interesting and beautiful'. A shape can be saved to a gallery by pressing the gray square at the top-right corner of the screen. The search-space (connected shapes only) holds ~36K shapes, with known distances between them. B) Gallery shapes were segmented into exploration and exploitation phases by the frequency of gallery shape choices. In exploitation player quickly scavenge a cluster of similar (visually or semantically related) shapes. In exploration players move in meandering paths, much longer than the shortest path between gallery shapes. Upon entering exploitation, they start to move in a directed way, following mostly shortest paths. Legend: circles = moves; colored circles = gallery-chosen shapes; gray = actual path; red = minimal paths. C) Network of clusters found by different players. Links connect clusters that share at least two shapes. Common meaning themes are defined as modules in the graph, and are shown in colors with representative shapes. D) Clusters of 'digits' found by different players. Transition shapes (left-most in each row) are the first shape in the cluster. Transition shapes are unique (eg the 'almost four' in the second row), while shapes that follow tend to be more prototypical in the sense that they are found also by the other players (e.g. the red highlighted '9' shape). E) A shape is ambiguous if it lies in the intersection of two contexts of meaning (e.g. the trident shape). Transition shapes are more likely to be ambiguous than exploitation shapes.

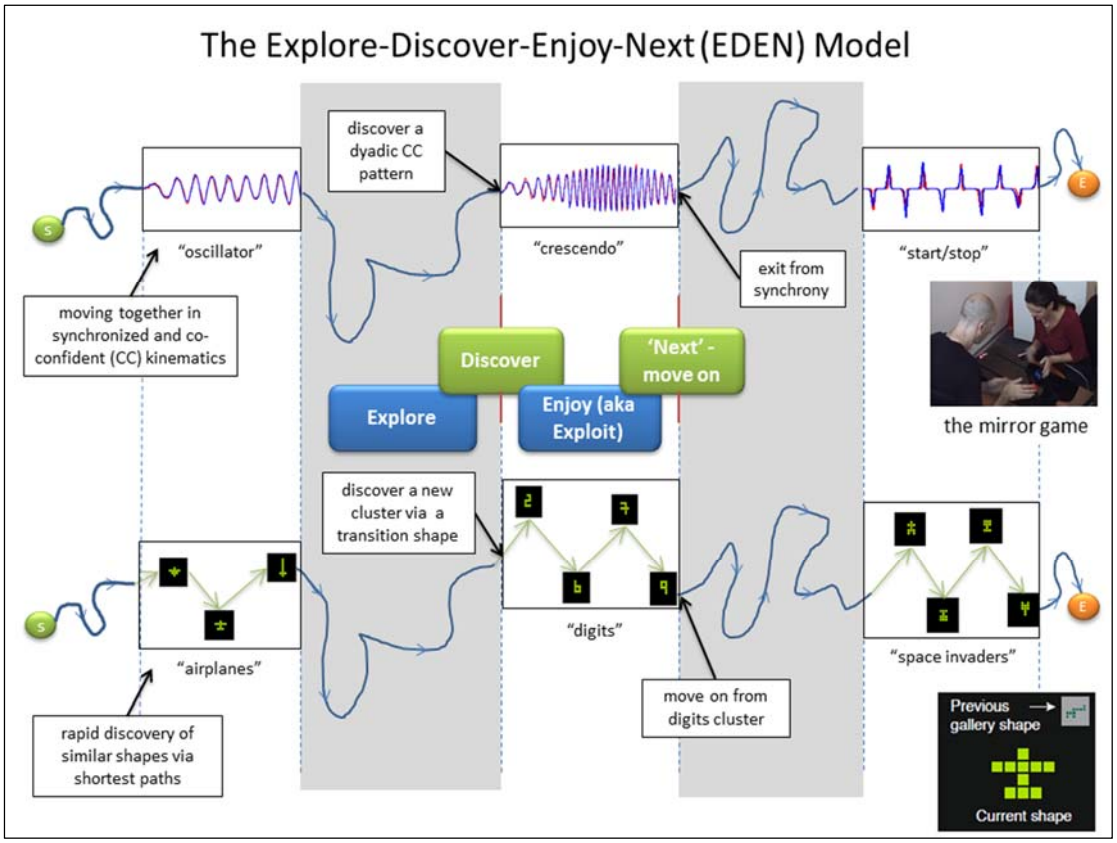

Figure 3. **A unifying framework for the two paradigms.** A schematic description of the dynamics in the MG (top) and in the CFG (bottom) paradigms. In both cases, the explorer (a dyad or an individual) move from left, starting at (s), and going through a sequence of alternating exploration and exploitations phases. Exploitation phases in the MG are periods were both players move in a synchronized and jitter-less motion (CC motion), according to an implicit agreement on the current pattern ("oscillator", "crescendo" or "start/stop" in this example). Exploitation phases in the CFG are clusters of similar shapes that are foraged quickly, along shortest path ("airplanes", "digits" or "space invaders" in this example). Exploitation phases start with a discovery point, and are terminated at some point by a decision to move on to the next exploration phase. In some case the 'next' decision stems from the dynamics of the exploitation phase – for example, in the MG, after a crescendo to the maximal possible frequency dyads often exit from synchrony either to anti-phase or to chaotic dynamics. The Explore-Discover-Enjoy-Next model extends current thinking on exploration/exploitation in two ways: highlighting the dynamical aspect of this process (going out of sync, and then back again), and focusing the research to the often neglected 'move-on' point. The processes underlying the decision to leave a region that still contains reasonable products, and going to explore the world again, are not known. Think for example on yourself as a scientist – when and how do you decide to leave a still productive line of research, and step into the unknown again, searching for more interesting discoveries?

References

Dahan, A., Noy, L., Hart, Y., Mayo, A. E., & Alon, U. (2016). Exit from synchrony in joint improvised motion. *PLOS ONE*, *11*(10), e0160747. http://doi.org/10.1371/journal.pone.0160747

Drucker, P. (2014). *Innovation and entrepreneurship*. Routledge.

Edersheim, E. H., & Drucker, P. F. (Peter F. (2007). *The definitive Drucker*. McGraw-Hill.

Feniger-Schaal, R., Noy, L., Hart, Y., Koren-Karie, N., Mayo, A. E., & Alon, U. (2016). Would you like to play together? Adults' attachment and the mirror game. *Attachment & Human Development*, *18*(1), 33–45.

Hart, Y., Mayo, A. E., Mayo, R., Rozenkrantz, L., Tendler, A., Alon, U., & Noy, L. (2017). Creative Foraging: an Experimental Paradigm for Studying Exploration and Discovery (under review). *PLoS One*.

Hart, Y., Noy, L., Feniger-Schaal, R., Mayo, A. E., & Alon, U. (2014). Individuality and togetherness in joint improvised motion. *PLOS ONE*, *9*(2), 1–8. http://doi.org/10.1371/journal.pone.0087213

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*(2), 431–40. http://doi.org/10.1037/a0027373

Hills, T. T., Todd, P. M., & Goldstone, R. L. (2015). The central executive as a search process: priming exploration and exploitation across domains. *Journal of Experimental Psychology. General*, *139*(4), 590–609. http://doi.org/10.1037/a0020666

Kneebone, R. (2011). The art , science , and simulation of performance. In *International Symposium on Performance Science.*

Kozbelt, A. (2006). Dynamic Evaluation of Matisse's 1935 Large reclining Nude. *Empirical Studies of the Arts*, *24*(2), 119–137. http://doi.org/10.2190/A2VY-TEBW-VH45-285E

Mendonca, D., Beroggi, G. E. G., & Wallace, W. A. (2001). Decision support for improvisation during emergency response operations. *International Journal of Emergency Management*, *1*(1), 30. http://doi.org/10.1504/IJEM.2001.000507

Noy, L. (2014). The mirror game : a natural science study of togetherness. In D. Citron, A., Aronson-Lehavi, S., & Zerbib (Ed.), *Performance Studies in Motion: International Perspectives and Practices in the Twenty-First Century.* (pp. 318–327). Bloomsbury Methuen Drama, London.

Noy, L., Alon, U., & Friedman, J. (2015). Corrective jitter motion shows similar individual frequencies for the arm and the finger. *Experimental Brain Research*, *233*(4), 1307–1320. http://doi.org/10.1007/s00221-015-4204-1

Noy, L., Dekel, E., & Alon, U. (2011). The mirror game as a paradigm for studying the dynamics of two people improvising motion together. *Proceedings of the National Academy of Sciences*, *108*(52), 20947–52. http://doi.org/10.1073/pnas.1108155108

Noy, L., Hart, Y., Andrew, N., Ramote, O., Mayo, A., & Alon, U. (2012). A quantitative study of creative leaps. In *Proc. Third Int. Conf. Comput. Creativity, Dublin* (pp. 72–76). Retrieved from http://computationalcreativity.net/iccc2012/wp-content/uploads/2012/05/072-Noy.pdf

Noy, L., Levit-Binun, N., & Golland, Y. (2015). Being in the zone: physiological markers of togetherness in joint improvisation. *Frontiers in Human Neuroscience*, *9*(187), 187. http://doi.org/10.3389/fnhum.2015.00187

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: a review of recent research (2006-2012). *Psychonomic Bulletin & Review*, *20*(3), 403–52. http://doi.org/10.3758/s13423-012-0371-2

Simonton, D. K. (1999). *Origins of genius: Darwinian perspectives on creativity*. New York, NY, US: Oxford University Press.

Weir, D. J., Stein, J. F., & Miall, R. C. (1989). Cues and control strategies in visually guided tracking. *Journal of Motor Behavior*, *21*(3), 185–204.