

Psychological perspectives on algorithm-based personnel selection

Cornelius J. König & Markus Langer
Saarland University
Saarbrücken, Germany



Co-funded by the
Erasmus+ Programme
of the European Union



14th Dutch-Flemish
Research Meeting on
Personnel Selection and Recruitment

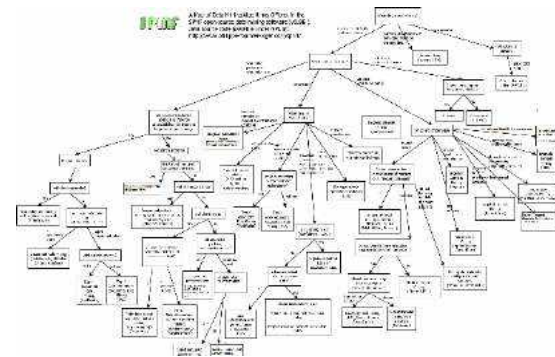
■ Automatic recognition of social behavior

- Sensors
 - Cameras
 - Microphones
 - Wearables



■ Automatic extraction of social behavior

- Machine learning algorithms



Video



Analyzer

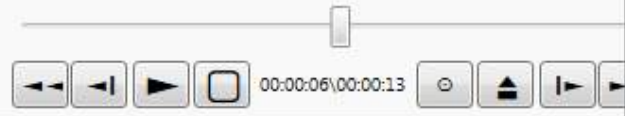
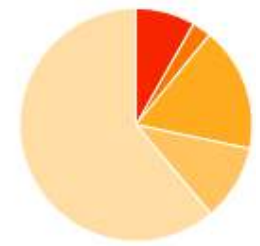
- Posture Time Abs
- Posture Time Rel
- HandsHeight
- Energy
- Fluidity
- Spat.Ext.
- Ovr. Act.
- En Gestures
- Fl Gestures
- Sp Gestures
- OA Gestures

Audio: VoiceActivity

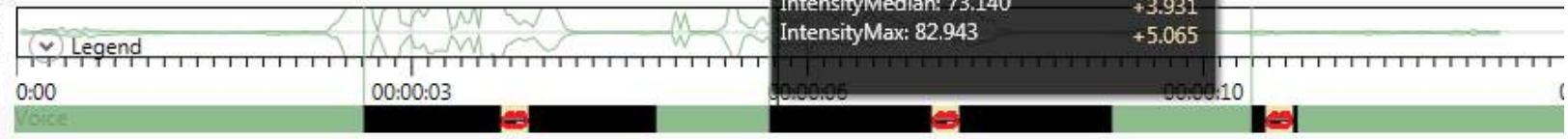


Duration: 2990 MilliSeconds
Begin: 6580 ms End: 9570 ms

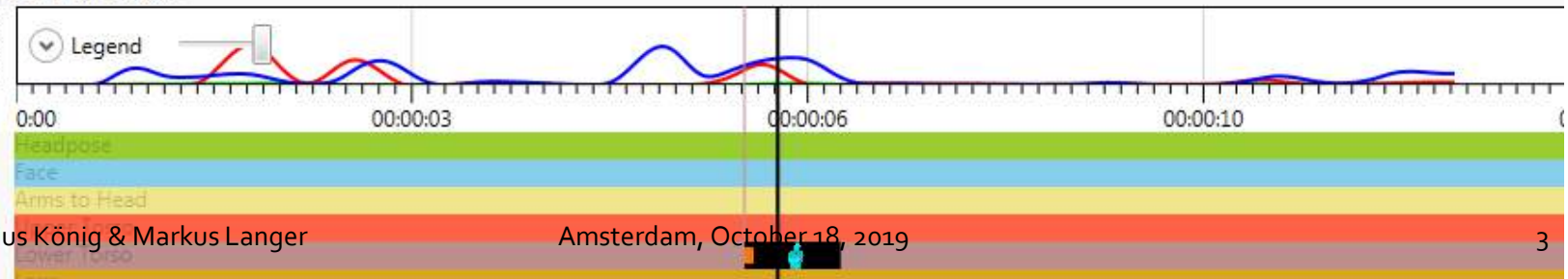
PitchMean: 102.048	+1.090
Pitch sd (Hz): 11.147	-27.607
Speechrate: 3.679	+0.443
Pulses: 50.167	+3.576
Jitter: 2.424	-1.850
Shimmer: 8.533	-2.800
Periods: 0.010	+0.000
Unvoiced Frames: 35.811	-3.693
Voicebreaks: 26.263	-7.089
Hamonicity: 0.854	+0.072
IntensityMedian: 73.140	+3.931
IntensityMax: 82.943	+5.065



Voice



Gestures and Postures



Social sensing in practice

- PRECIRE
 - Screening for personnel selection
 - Training
- HireVue
 - Screening for personnel selection
 - Training

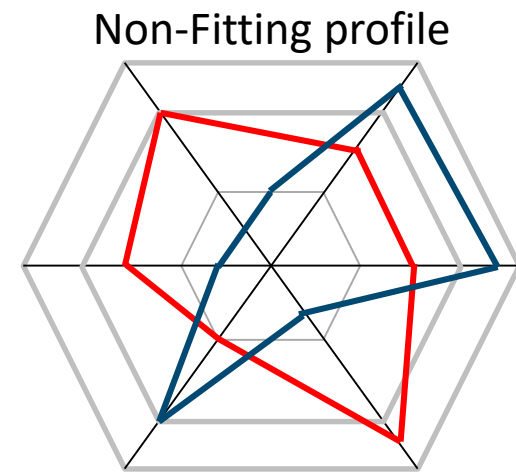
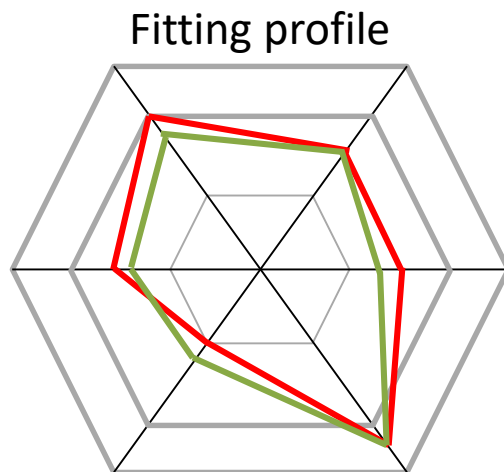
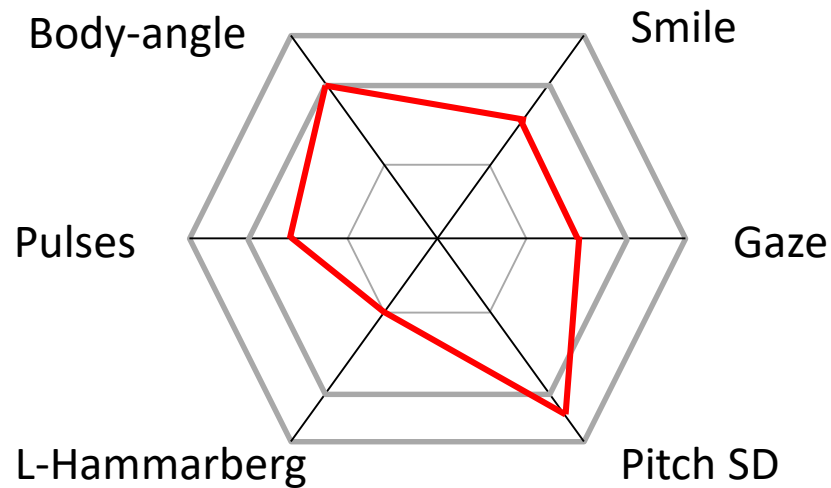


PRECIRE
TECHNOLOGIES



HireVue
AMAZING DIGITAL INTERVIEWS

Profile comparisons



- Focus on applicants
 - Do applicants behave differently when experiencing algorithm-based interviews?
 - How do applicants react to algorithm-based interviews?
 - Are algorithm-based interviews valid?
- Focus on recruiters

Do applicants behave differently?

Langer, M., König, C. J., & Hemsing, V. (re-submitted). *Is anybody listening? The impact of automatically evaluated job interviews on impression management and applicant reactions.*

- Main questions: If applicants know that their interview responses are automatically assessed by some algorithms, does this change their behavior?

Do applicants behave differently? (II)

- Main hypothesis (based on theorizing by Levashina & Campion, 2006, and by Marcus, 2009):
 - *Participants in an automatic evaluation condition will report using less honest and deceptive IM behavior than participants in a human rater condition.*
- Online study with $N = 124$ (mainly students)
- Hypothetical asynchronous interview

Do applicants behave differently? (III)

- Manipulation:
 - The human rater group was told that *“a member of the department of industrial and organizational psychology with experience in personnel selection will listen to the audio recordings and evaluate your answers.”*
 - The automatic evaluation group was told that *“a computer will automatically analyze the audio recordings and evaluate your answers.”*
- Four interview questions
 - No recording (for technical reasons)

Do applicants behave differently? (IV)

- Main dependent variables
 - 5 items honest IM
 - 6 items deceptive IM (both from Roulin & Bourdage, 2017)



Do applicants behave differently? (V)

■ Results

- Significantly less deception IM, $d = -0.35^*$
- But similar amount of honest IM, $d = -0.13^{n.s.}$
- Additional result
 - Participants in the automatic evaluation condition spent significantly less time on these pages, $d = -0.41^*$
 - Automatic evaluation $M = 65.62$ s, human rater $M = 84.04$ s

■ Conclusion

- Yes, applicants behave differently!

How do applicants react to algorithm-based interviews?

Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly-automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, 27, 271-234.

■ Main questions

- How do applicant react to algorithm-based interviews?
- And do they take the context into account?
 - High stakes (selection) vs. low stakes (training)

How do applicants react to algorithm-based interviews? (II)

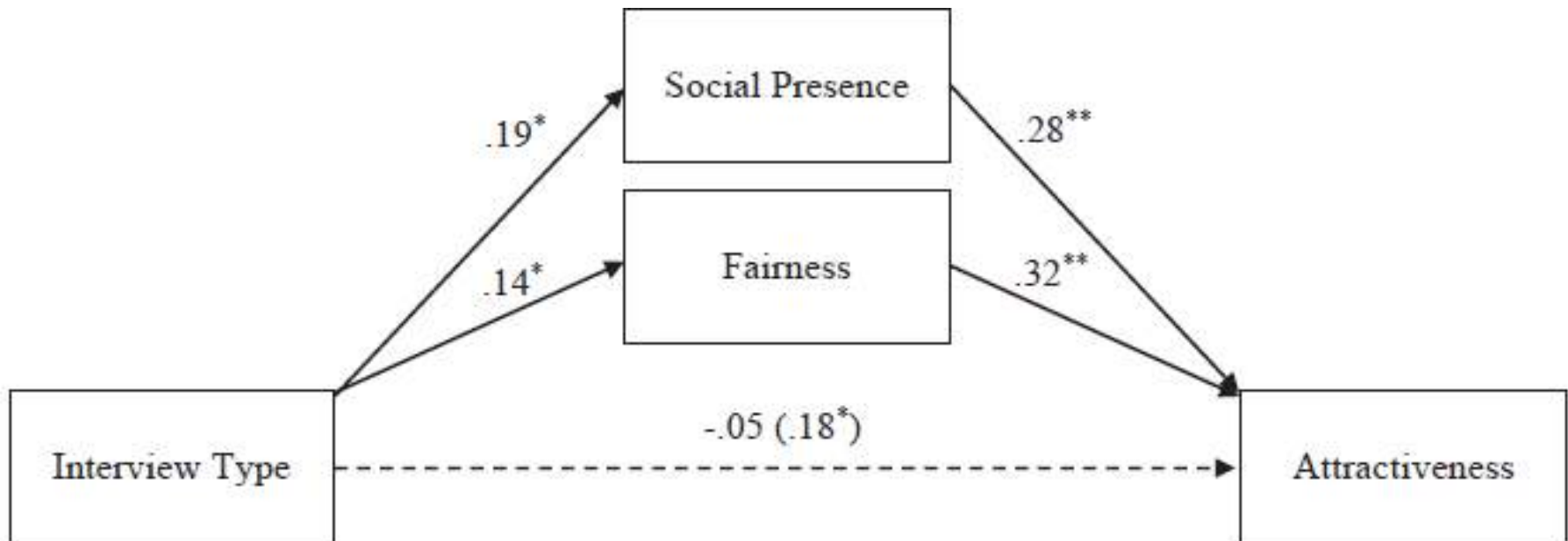
■ Methods

- Online study, $N = 123$ (mainly students)
- 2×2 between subject design
 - Videoconference vs. algorithm-based interview (with an avatar)
 - Low-stakes (i.e., training) vs. high-stakes context (i.e., selection)
- Observers scenario
 - Participants were asked to imagine that a friend was invited to a job interview
- Dependent variables
 - Standard self-reports including fairness, creepiness



How do applicants react to algorithm-based interviews? (III)

- Main results



How do applicants react to algorithm-based interviews? (IV)

- Main results (cont.)
 - Negative effects slightly more pronounced in the selection (vs. training) condition
- Conclusions
 - Applicants will probably not like algorithm-based interviews that much

Can we increase applicant reactions by giving them more information?

Langer, M., König, C. J., & Fitali, A. (2018). Information as a double-edged sword: The role of computer experience and information on applicant reactions towards novel technologies for personnel selection. *Computers in Human Behavior, 81*, 19-30.

- New field within AI: eXplainable AI (XAI)
- What if we explain to applicants what these algorithms do?

Can we increase applicant reactions by giving them more information? (II)

■ Methods

- Online study, $N = 120$ students
- Hypothetical scenario
 - Interview with an avatar plus algorithm-based evaluation
- 2×2 between subject design
 - Low vs. high amount of information (→ next page)
 - Non-computer science vs. computer science students
 - [But no effects]
- Dependent variables
 - Standard self-reports including fairness, creepiness etc.



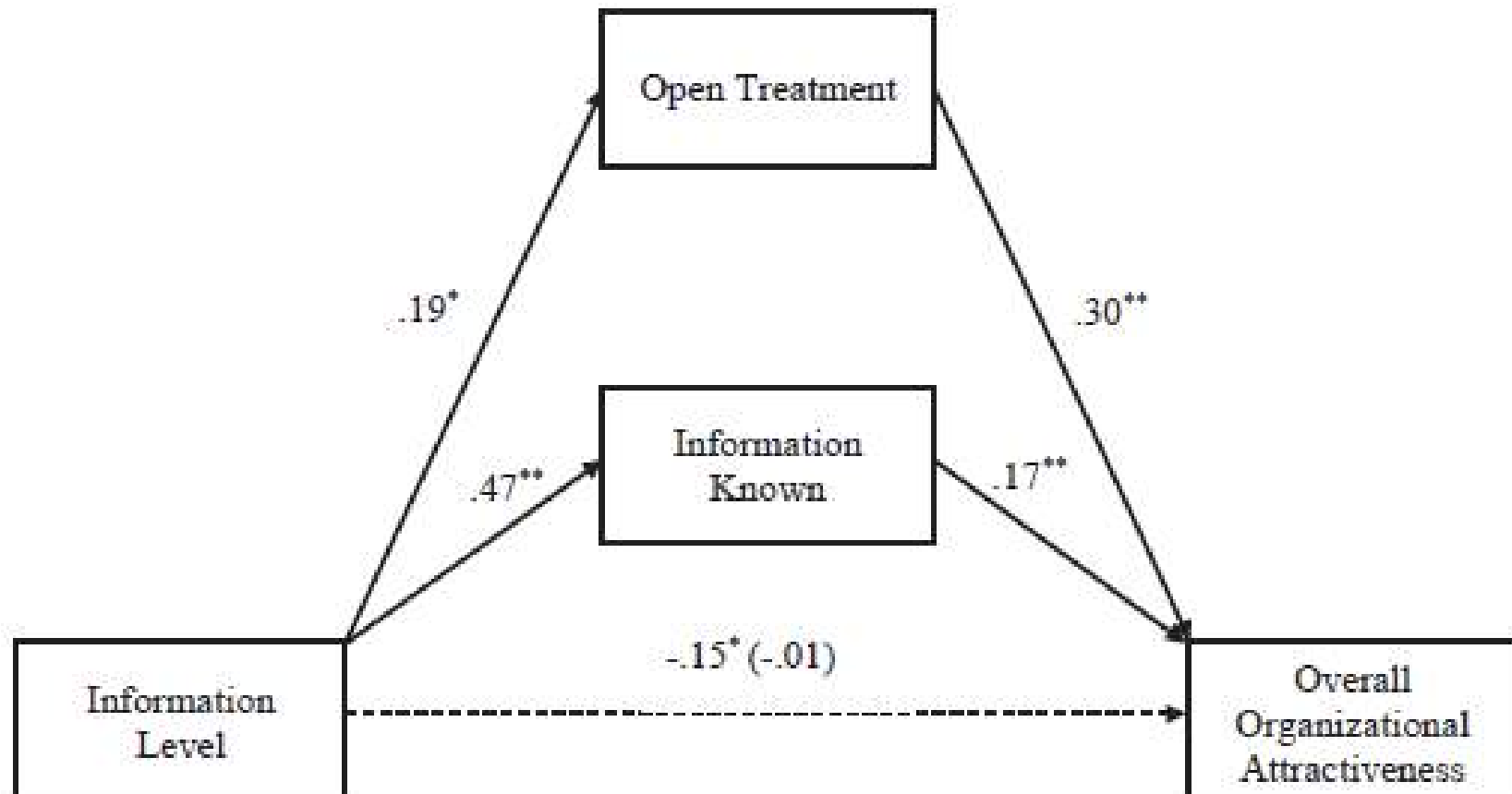
Can we increase applicant reactions by giving them more information? (III)

Main pieces of information given only to the high information group:

- “The program can ...
 - ... analyze your speech and voice pauses because such signals can be used to infer personality traits
 - ... express human communication aspects through the virtual character because studies showed that a virtual character with human communication aspects is perceived as more likable
 - ...analyze your facial expressions [because...]
 - ... analyze your gestures by recognizing hand, body, and head movement [because...]
 - ... interpret your behavior as social and emotional signals [because...]
 - ... adapt to your individual behavior [because...]”

Can we increase applicant reactions by giving them more information? (IV)

- Main result (a suppressor effect):



Can different kind of information help?

Langer, M., Baum, K., König, C. J., & Hähne, V. (submitted).

- Idea: two different aspects of information
 - Process information
 - i.e., what the computer program can do
 - Justification information
 - i.e., why the computer program automatically analyzes applicants
- Study design
 - 2 × 2 between subject design
(process information: yes vs. no;
justification information: yes vs. no)



Can different kind of information help? (II)

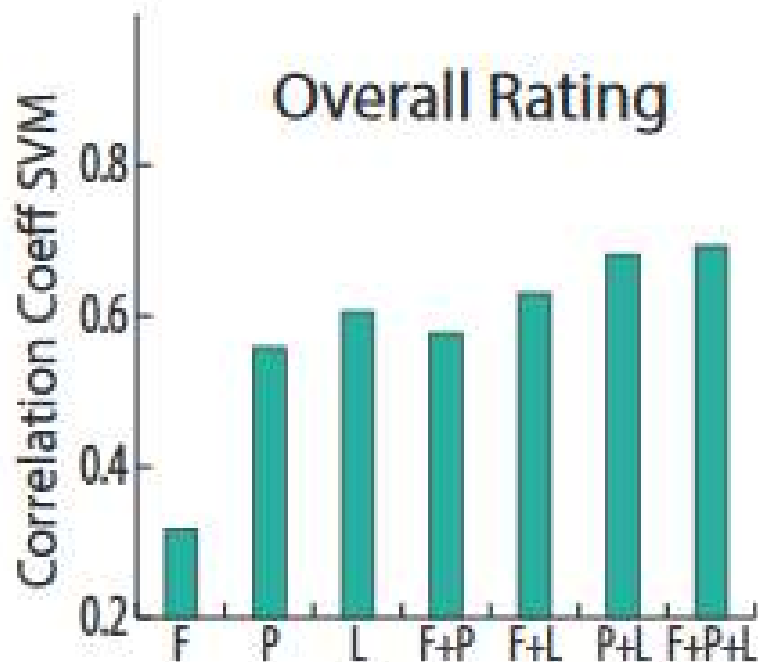
- Main results
 - Process information can induce negative emotional reactions and increase privacy concerns
 - Justification information can increase perceived fairness
 - Unexpectedly,
 - ... providing no information may not be that detrimental
 - ... when information is presented, perceived transparency does not necessarily increase
- Conclusions
 - Organization should use justification information – or no information (?)

Are algorithm-based interviews valid?

- Only Naim et al. (2018) (plus Schmid Mast et al., under review)
 - Naim, I., Tanveer, M. I., Gildea, D., & Hoque, M. E. (2018). Automated analysis and prediction of job interview performance. *IEEE Transactions on Affective Computing*, 9, 191–204.
[doi:10.1109/TAFFC.2016.2614299](https://doi.org/10.1109/TAFFC.2016.2614299)
 - Database: 69 students with 2 short mock interviews, each with 1 professional career counselor (videos plus transcriptions made by MTurkers)
 - Criterion: interview ratings by 9 Mturkers
 - Algorithm-based evaluation of the interviews based on facial, prosodic (speech), and lexical (words) features
 - Statistical approach: Support Vector Machine and Lasso regressions

Are algorithm-based interviews valid? (II)

- Results of Naim et al. (2018)



F = Facial features
P = prosodic features
L = lexical features

Are algorithm-based interviews valid? (III)

- Why not more research?
 - Many W/O psychologist are very, very skeptical...
 - Knowledge problem? → EU project Big Data in Psychological Assessment (BDPA, see <https://bdpa.eu>), offering free teaching material



- There are more stakeholders than applicants!
 - Recruiters
 - Companies selling AI products
 - Work councils, unions, NGOs etc.
- Particularly relevant if something goes wrong: Who is responsible?
- 2 studies
 - How does automated decision-support affect the work of recruiters?
 - How do recruiters react if pre-selection by an AI system is unfair?

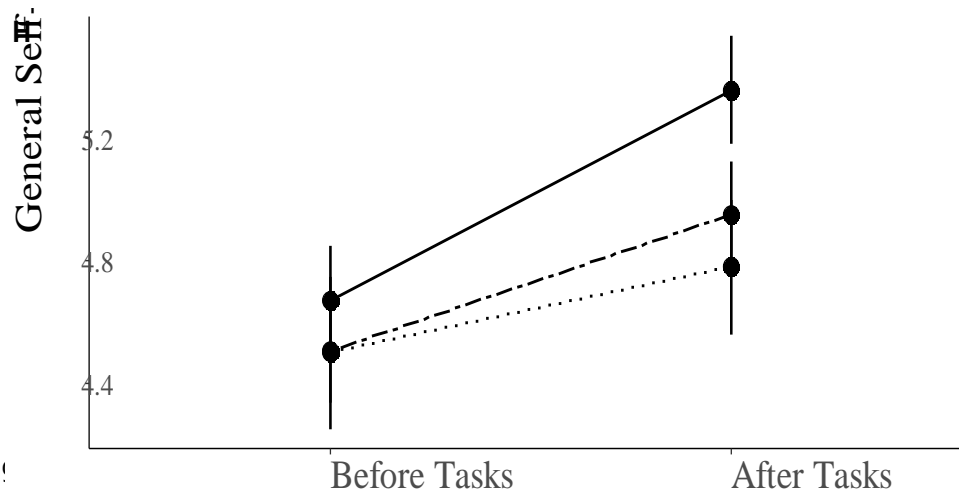
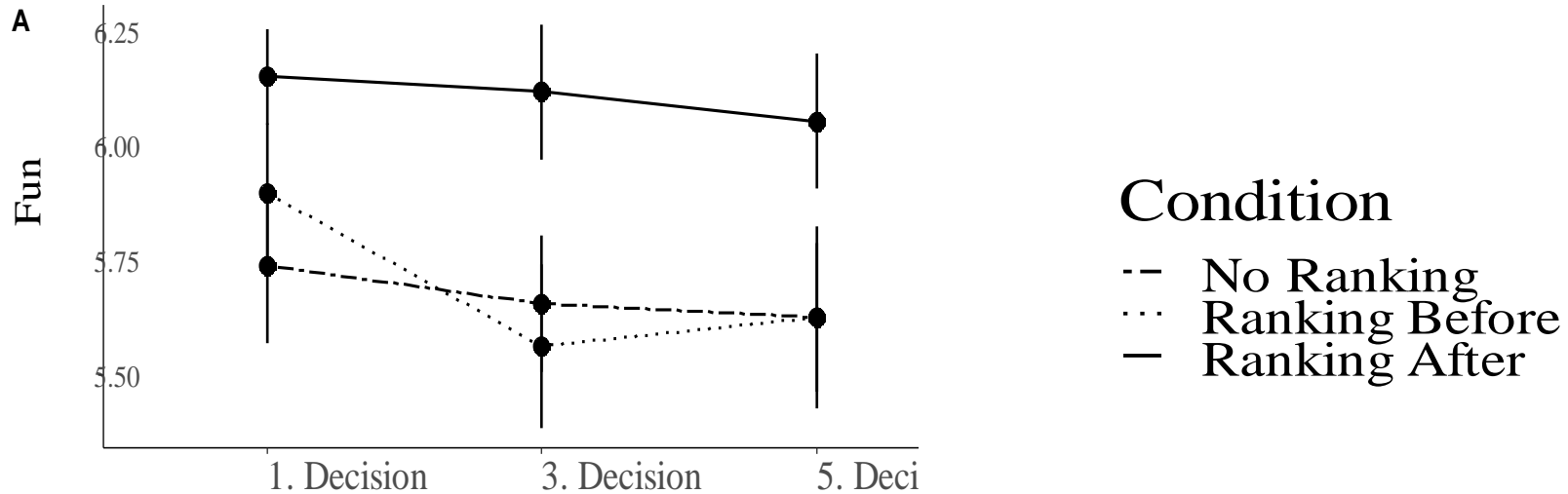
How does automated decision-support affect recruiters?

Langer, M., König, C. J., & Busch, V. (in preparation).

- Future scenario: Recruiters will be provided with candidate lists determined by AI
- What does this mean for the work experience of recruiters?
- Lab study ($N = 81$ students)
 - 5 hiring tasks (choose 1 out of 6 candidates based on self-presentation audio files)
 - 3 groups: no-ranking vs. ranking-before-the-decision. ranking-after-the-decision



How does automated decision-support affect recruiters? (II)



How does automated decision-support affect recruiters? (III)

- Conclusion:
 - Decision-support systems can affect fun and task satisfaction
 - AI-based recommendation as feedback and an opportunity to reflect on their decision

How do recruiters react if pre-selection by AI is unfair?

Feldkamp, T., Langer, M., König, C. J., & Wies, L. (in preparation).



Balanced pre-selection

 <p>Piaula Koss Berufserfahrung: 9 Jahre Stärken: Selbstorganisation, Priorisierung Schwächen: Durchsetzungsvorgänge Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>Till Becker Berufserfahrung: 4 Jahre Stärken: Leistung unter Stress und Zeitdruck Schwächen: Schwache Konfliktkompetenz Durchschnittliche Leistung der Arbeitszeugnisse: 1,2</p>	 <p>Evelyn Laska Berufserfahrung: 4 Jahre Stärken: Kommunikation und Kooperationsfähigkeit Schwächen: Selbstorganisation Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>Maurice Kopp Berufserfahrung: 7 Jahre Stärken: Priorisierung, Teamfähigkeit Schwächen: Mangelndes Zeitmanagement Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>
 <p>Benedikt Helmert Berufserfahrung: 2 Jahre Stärken: Schnelle Anpassung an neue Themen Schwächen: Konzentration unter Stress Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>Eike Kell Berufserfahrung: 2 Jahre Stärken: Belastbarkeit unter Stress Schwächen: wenig Priorisierung Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>Lukas Dreckert Berufserfahrung: 2 Jahre Stärken: Kreativität, Durchsetzungsvermögen Schwächen: Geringe Priorisierung Durchschnittliche Leistung der Arbeitszeugnisse: 1,2</p>	 <p>Milena Seiser Berufserfahrung: 1 Jahre Stärken: Zielvorgabe, Offenheit Schwächen: Verhalten in großen Gruppen Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>
 <p>Anna Kasper Berufserfahrung: 7 Jahre Stärken: Teamfähigkeit, Kreativität Schwächen: Überforderung unter Stress Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>Anja Rühl Berufserfahrung: 4 Jahre Stärken: Gutes Zeitmanagement Schwächen: Konfliktkompetenz Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>Marco Babenhörl Berufserfahrung: 2 Jahre Stärken: Teamfähigkeit, Zeitmanagement Schwächen: Fähigkeit zum sozialen Handeln Durchschnittliche Leistung der Arbeitszeugnisse: 1,4</p>	 <p>Lukas Winkler Berufserfahrung: 7 Jahre Stärken: Praxiserfahrung im Bereich Personal Schwächen: Selbstorganisation Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>

Unbalanced pre-selection

 <p>Simon Böler Berufserfahrung: 2 Jahre Stärken: Sehr gute Lernfähigkeit, Wissensorientiert Schwächen: Geringe Praxiserfahrung Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>Till Berber Berufserfahrung: 4 Jahre Stärken: Leistung unter Stress und Zeitdruck Schwächen: Schwacher Konfliktlösungskompetenz Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>Robert Jahn Berufserfahrung: 5 Jahre Stärken: Starke Durchsetzungsvermögen Schwächen: Nervosität bei neuen Aufgaben Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>Maurice Kapp Berufserfahrung: 7 Jahre Stärken: Praxiserfahrung, Teamfähigkeit Schwächen: Mäßiges Zeitmanagement Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>
 <p>Anni Blatt Berufserfahrung: 4 Jahre Stärken: Gutes Zeitmanagement Schwächen: Konzentrationsprobleme Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>Cedric Mehnert Berufserfahrung: 2 Jahre Stärken: Sehr gute Selbstorganisation Schwächen: Geringe Praxiserfahrung Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>David Kemper Berufserfahrung: 3 Jahre Stärken: Gute Teamfähigkeit, Zeitmanagement Schwächen: Durchsetzungsvermögen Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>Eric Glöckner Berufserfahrung: 3 Jahre Stärken: Kreativität, innovative Ideen Schwächen: Bessert sich viel Ruhe zum Arbeiten Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>
 <p>Lukas Decker Berufserfahrung: 2 Jahre Stärken: Kreativität, Durchsetzungsvermögen Schwächen: Geringe Praxiserfahrung Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>Britta Kahl Berufserfahrung: 2 Jahre Stärken: Belastbarkeit unter Stress Schwächen: Wenig Praxiserfahrung Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>	 <p>Marcel Eichenkorn Berufserfahrung: 5 Jahre Stärken: Teamfähigkeit, Zeitmanagement Schwächen: Flexibilität kann zu wenig Durchschnittliche Leistung der Arbeitszeugnisse: 1,4</p>	 <p>Linus Weiden Berufserfahrung: 7 Jahre Stärken: Praxiserfahrung im Bereich Personal Schwächen: Selbstorganisation Durchschnittliche Leistung der Arbeitszeugnisse: 1,3</p>

How do recruiters react if pre-selection by AI is unfair? (IV)

- Design: 2 (human vs. algorithm-based recommendation) × 2 (balanced vs. unbalanced preselection) between-subject experiment
- Main results:
 - A balanced preselection was perceived as fairer, but reliance on the preselection was unaffected
 - Human recommendation was perceived as more biased by prejudices
 - Likely different moral judgments (maximization of utility vs. “doing it the right way” – but we haven’t finished analyzing the qualitative data)

- Research from our lab
 - Results from studies on traditional interviews will not automatically generalize to algorithm-based interviews
 - Applicants react predominantly negative to them
 - Despite the EU's General Data Protection Regulation and the XAI hype: Finding appropriate ways to inform applicants will be difficult
 - AI recommender systems will affect recruiters' work
- Not enough research!
 - In particular on the validity of algorithm-based selection procedures
 - And on other stakeholders

How could the field proceed?

- Challenges we need to tackle
 - Combining human factors and personnel selection literatures
 - e.g., trust in automation
 - More collaboration with colleagues from other fields
 - In particular computer scientists and philosophers
 - Showing the relevance of W/O psychology research to the public

Thank you for listening

