Measuring personality in the presence of faking

Anna Brown

Honorary Professor, University of Kent

European Conference on Personality, Berlin, 2024

Why measure personality?

- Personality predicts important life outcomes
  - Mortality, divorce, occupational attainment (Roberts et al., 2006)
  - Subjective and psychological well-being (Anglim et al., 2022)
  - Higher grades, better job performance in entry-level jobs (Bakker et al., 2012), early career outcomes (Hoff et al., 2021)
  - Job performance (Rothmann & Coetzer, 2003; Dudley et al., 2006); job satisfaction (Allemand et al., 2015), work engagement (Bakker et al., 2012)
  - Health status, substance use, anxiety and depression (Wright & Jackson, 2022)

- Personality assessment should be a valuable tool
  - Basic and applied research
  - Great interest in assessing personality in job applications, educational admissions and other high-stakes contexts (Ryan et al., 2015)

Measuring personality in high stakes

- Personality is almost exclusively assessed via self-reports

		Strongly Disagree	Disagree	Unsure	Agree	Strongly Agree
1	I enjoy meeting new people.	○	○	○	◉	○
2	I like helping people.	○	○	○	○	○
3	I sometimes make mistakes.	○	○	○	○	○
4	I'm easily disappointed.	○	○	○	○	○

- Candidates can easily manipulate their responses

Faking

- Intentional (motivated) distortion of the true self-image
  - with the goal of creating a desired impression (i.e. impression management)
    - favourable impression ("faking good")
    - unfavourable impression ("faking bad")

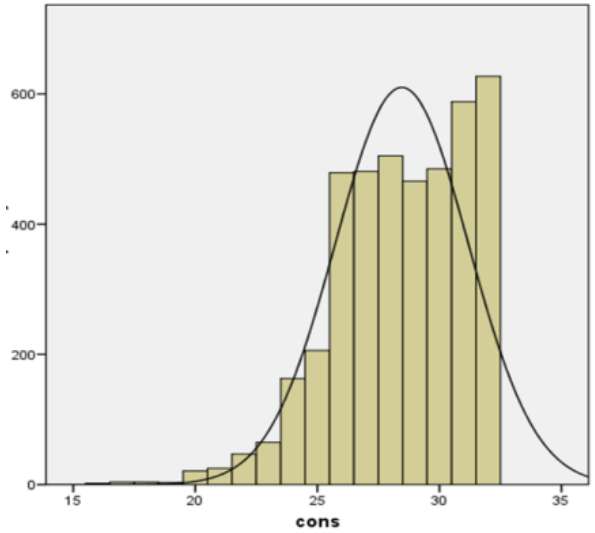
- Do people fake?
- If so, does it matter?
- If so, what can we do about it?

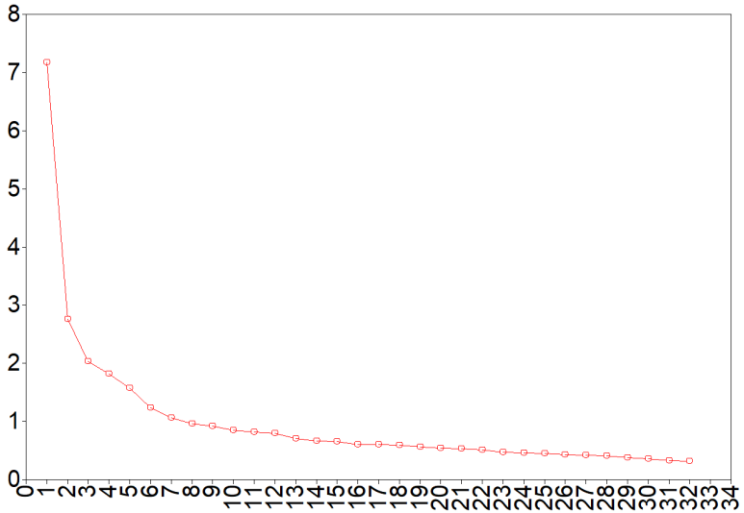
Do people fake?

- "*In 2020, individual assessors do not believe that a large number of candidates fake, and believe that even fewer candidates successfully fake*" (Robie et al., 2021)

- People do fake. Estimated 30-50% of job applicants engage in faking to pursue hiring success
  - People admit to faking (e.g., König et al., 2011)
  - We know that people fake from within-participants research designs (Donovan et al., 2014; Griffith & Converse, 2011; Peterson et al., 2011)
    - Applicants score higher than incumbents and research participants on all Big 5 dimensions except Openness (Birkeland et al., 2006)
    - Faking is not a personality trait!
  - We know that people fake from psychometric evidence

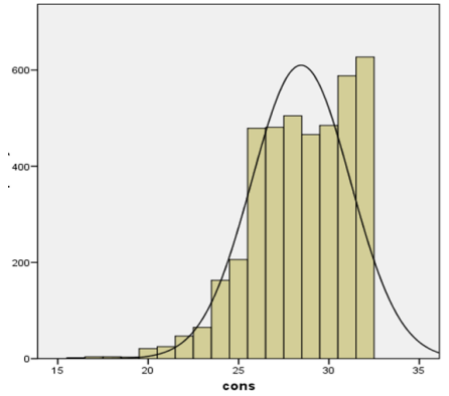
Psychometric evidence of faking





- Trait score distributions are skewed
  - Faking changes the true ordering of people!
- All trait scores correlate with each other, regardless of their conceptual concordance
  - Controversy of the General Factor of Personality (Pelt, 2019)
- Cronbach's alpha of faked scores is high
  - Faking is not a type of “inconsistent” or “aberrant” responding!

Does faking matter?



- One point of view is that faking **does not matter**
  - Faking is *"saying what you think you ought to say rather than what you really want to say. We have a word for that – "civilization."* (Kevin Murphy, in Morgeson et al., 2007)
  - Test scores in high stakes still predict job performance (e.g. Iliescu et al., 2021)
    - employees continue "managing impression" after hire (Ones et al., 2007)
  - However, correlation with criterion is driven by the bottom part of the distribution (e.g. Mueller-Hanson et al., 2003)
- Another point of view is that faking **matters** (e.g. Brown & Böckenholt, 2022; Tett & Simonet, 2021)
  - It destroys the construct validity – because the test no longer measures personality but what we "think we are ought to say"
  - It compromises fairness, giving advantage to fakers

What can we do about faking?

- Detect faking after it has occurred, and adjust (correct) the test scores statistically
- Prevent faking before it has occurred
- Abandon self-reported measures of personality and invent something better

An alternative to personality self-reports

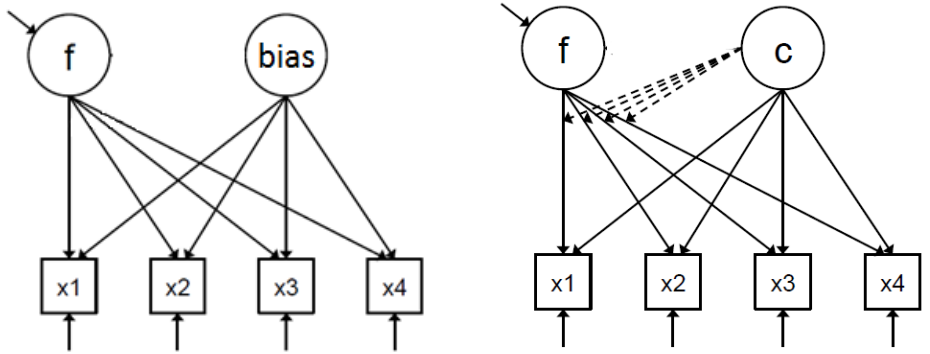
- If faking reflects the adherence to social norms, i.e. "civilization" (Murphy, 2007), why don’t we just ask the candidates:
  - What kind of person do you think we would like to recruit? (the “ideal-employee” image as they see it)
  - **and**, How motivated are you to get this job?

- Taken together, the ability to identify criteria (ATIC) and motivation explain a sizeable amount of variance in job performance
  - And there is little reason to fake the above measures

1. Detection and correction

What can we do about faking?

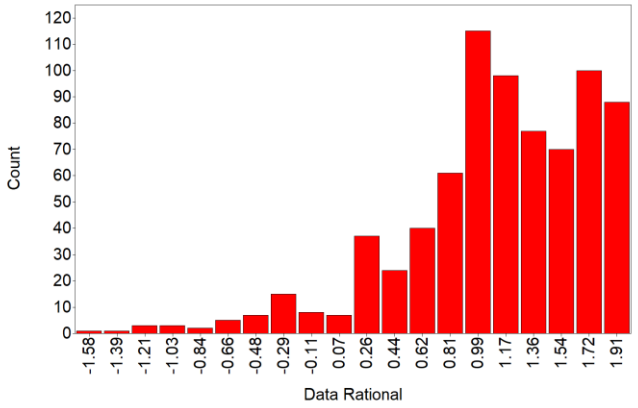
(Some) Detection and correction methods

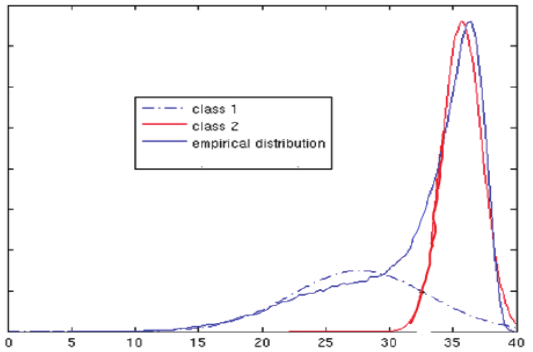


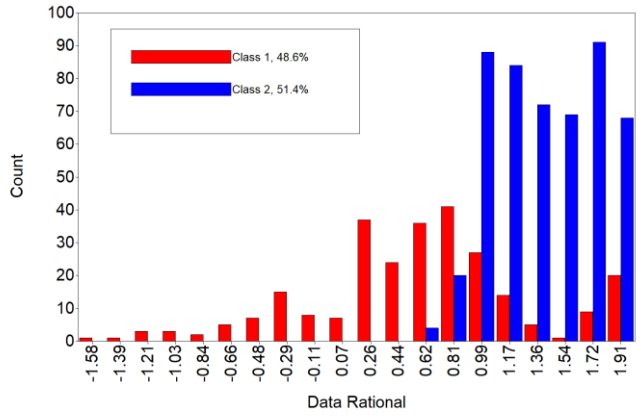
- Manifest / Observed variables
  - Social Desirability scales (unlikely virtues);
  - Overclaiming techniques (foil questions);
  - Blatant extreme responding (BER)
  - Problematic scores are either **flagged**, or used to **correct** the personality scores
    - E.g. the regression residual of trait score on the "social desirability index" is assumed free of bias (Webster, 1958)
- Latent variables
  - **Latent factors**, e.g. 'Ideal Employee Factor' (Klehe et al., 2011)
  - **Latent classes** - 'honest', 'mild fakers', 'extreme fakers' (Zickar, Gibby & Robie, 2004)
- Hybrids

Faking as a continuum

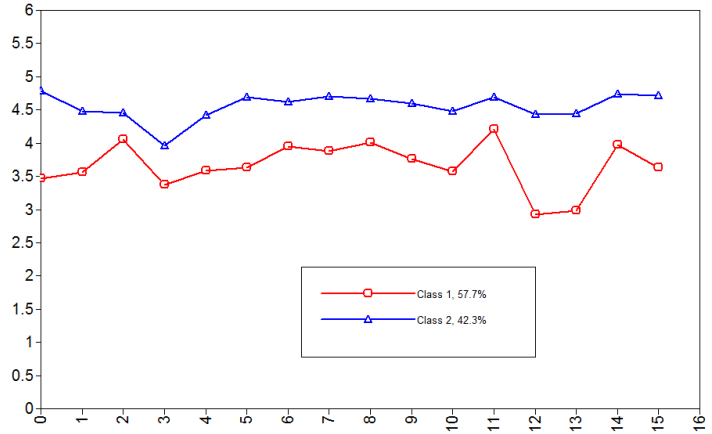
- Faking is a continuum, and everyone has a score according to their "extent of faking"
  - **Manifest indices** measure this extent directly
  - In **latent factor** models, all person responses indicate not only their dedicated trait(s), but also the person's standing on the "Ideal Employee Factor"
- But additive effects "true score + faking score" can **never** explain the skewed distributions we observe in high stakes (Brown & Böckenholt, 2022)
  - These distributions are only possible if there is heterogeneity in response mechanisms







Faking as a class



- Person is a member in one of several classes according to their "faking response behaviour"
- Re-analysis of Brown (2008) study
  - Instructed faking / Honest conditions
  - One job description was used as target; should yield the same "ideal" profile
  - Latent class analysis (LCA) of 16 personality trait scores
    - 2 classes give excellent separation (entropy = .984) of “real” (58%) and “ideal” profiles (42%)
  - Latent classes largely coincide with the experimental conditions

- Unfortunately, LCA does not achieve such neat results with operational data
  - 3 classes were needed to describe the data in Zickar, Gibby & Robie (2004)

Criticism of latent factor/class approaches

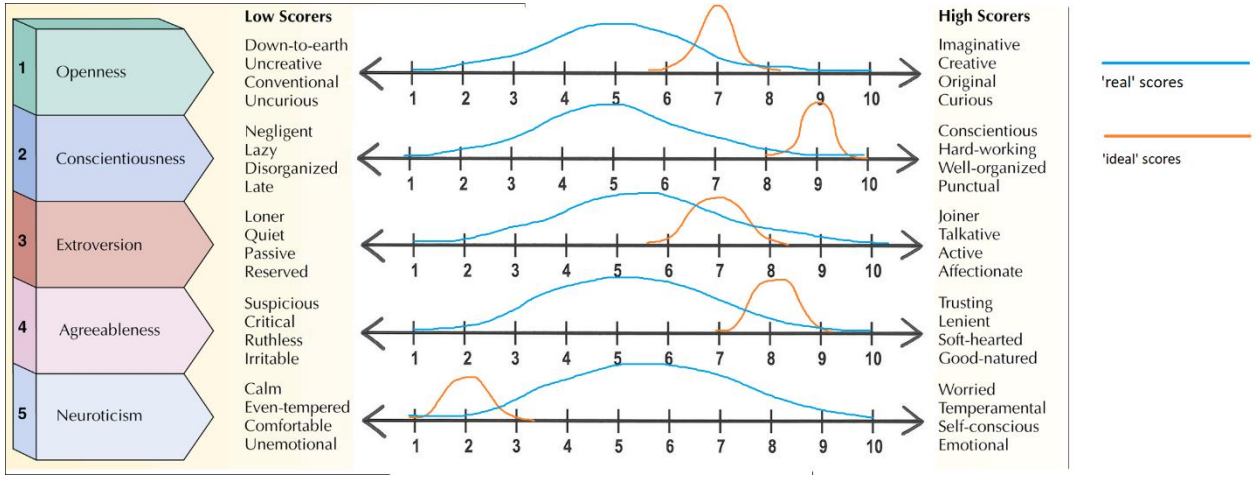
- Both approaches assume that faking is consistent across the questionnaire
  - Candidates fake **all items**, but differ in the **amount** they fake them
  - Each candidate has **only one position** of the “ideal employee” factor, expecting to fake all items to some extent
  - Each candidate is a member of **only one class** (e.g. honest or faking), so they are assumed to fake nothing or everything, or everything but only "mildly"
- Brown & Böckenholt (2022) argue that, due to the false assumption of consistent faking,
  - "Ideal Employee Factor" over-extracts useful variance
  - Intermediate classes are methodological artefacts
    - Why would candidates intentionally produce profiles that are only “half-good” to get a job?

Psychology of faking behaviour

- Kuncel & Tellegen (2009) researched test taker cognitions
  - Test takers are communicating with a prospective employer, and they want to be impressive, credible, and true to themselves while taking the test
  - These motives are potentially conflicting
- Ariely (2012) showed that people cheat only as much as they must, to maintain self-concept of honourable people
  - Faking happens on selected questions and attributes

- To overcome the limitations of the latent variable models, we need to consider intermittent faking – person can fake only some items but not others

Intermittent faking



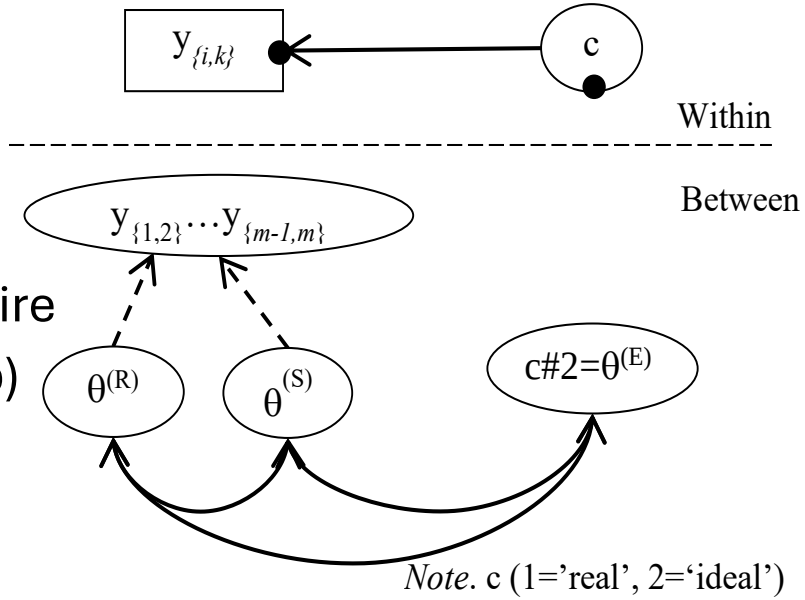
- There exist 2 profiles – 'real' and 'ideal' – each with their own distribution and factor structure
- Response of every person to every item may be from either profile
  - Candidate's profile is a mix of 'real' and 'ideal' responses
- Candidate can be a **partial member** in the two profiles
  - Unlike simple latent class models allowing only exclusive class membership, we consider "grades of membership"

(Brown, A., & Böckenholt, U. (2022). Intermittent Faking of Personality Profiles in High-Stakes Assessments: A Grade of Membership Analysis. Psychological Methods, 27(5), 895-916.

Faking as Grade of Membership model

- The F-GOM model is operationalised as a two-level mixture
  - 2 classes at the item level, and a factor structure at the person level

- F-GOM model can estimate:
  - Probability of each response to be 'real' or 'ideal'
  - Scores on the substantive traits measured by the questionnaire
  - The tendency to Edit one's responses (**grade of membership**)
  - The tendency to Select extreme responses when editing



Note. c (1='real', 2='ideal')

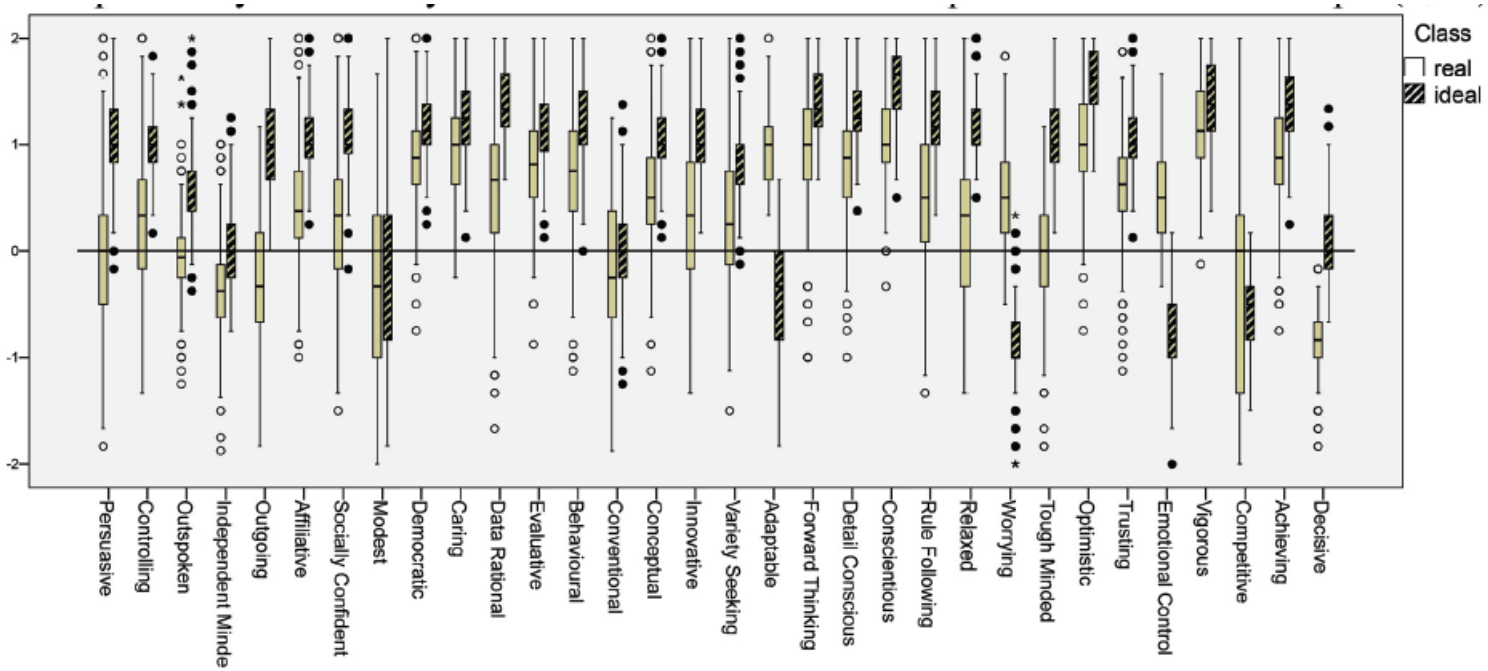
Grade-of-membership analysis in action

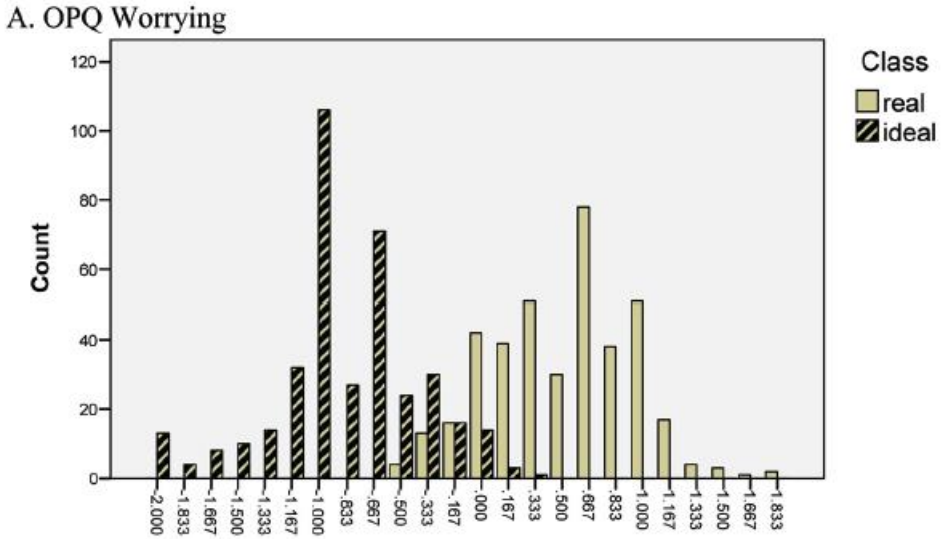
- Brown & Böckenholt (2022) analysed an archival dataset from a large USA retailer
- N = 762 applicants to analytical jobs (e.g., finance analyst)
- Available data
  - OPQ32n - 32 work-related personal styles
  - Social Desirability scale (“unlikely virtues”)
  - Ability tests (Verbal VC11; Numerical NC21)

- F-GoM model fitted the data much better than any of the alternatives (latent class and factor models)

GOM analysis: Distributions

- ‘Real’ profile (43% of all scores) and ‘ideal’ profile (57% of all scores) had different means and score distributions





GOM analysis: Score validity

- Scores classified as ‘real’ had higher validity than ‘ideal’
  - **Verbal** test correlated with OPQ **Evaluative** .35** in ‘real’ class but .07 in ‘ideal’ class
  - **Numerical** test correlated with OPQ **Data Rational** .40** in ‘real’ class but .09* in ‘ideal’ class

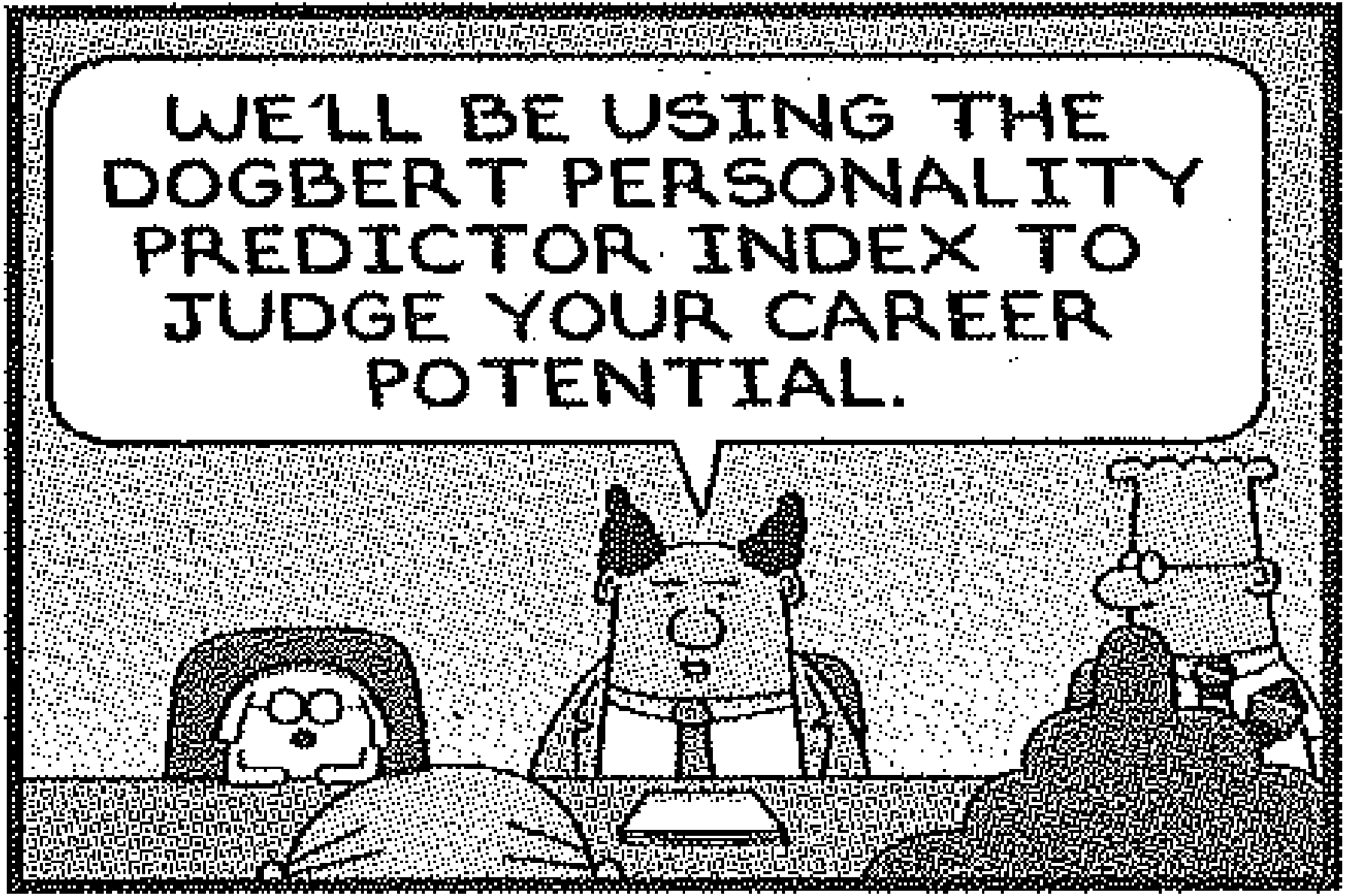
- OPQ **Social Desirability** scale correlated with OPQ substantive traits as well as tendencies to **Edit** and to **Select** extreme responses when editing
  - Social Desirability scales mix both substance and style!

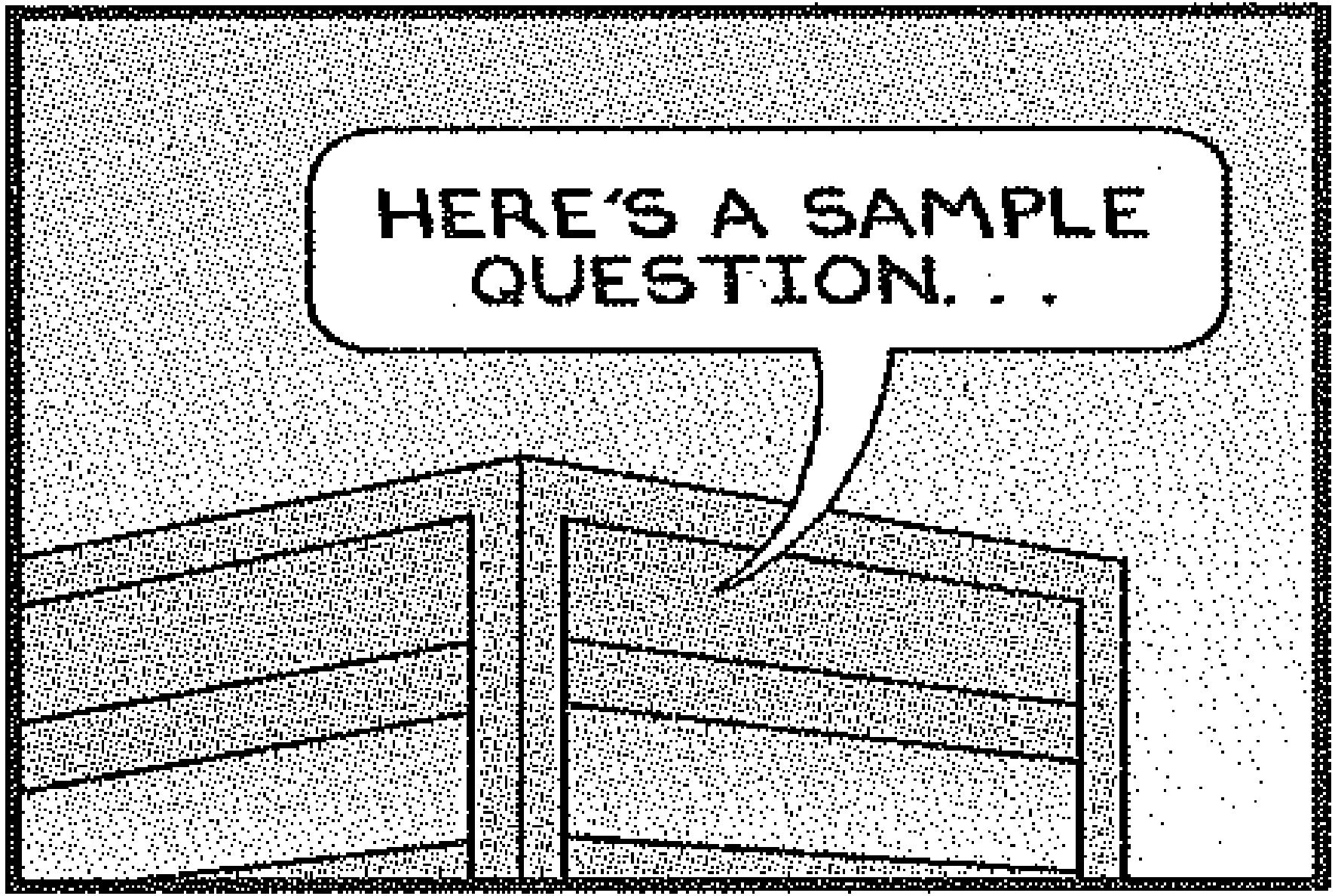
2. Prevention

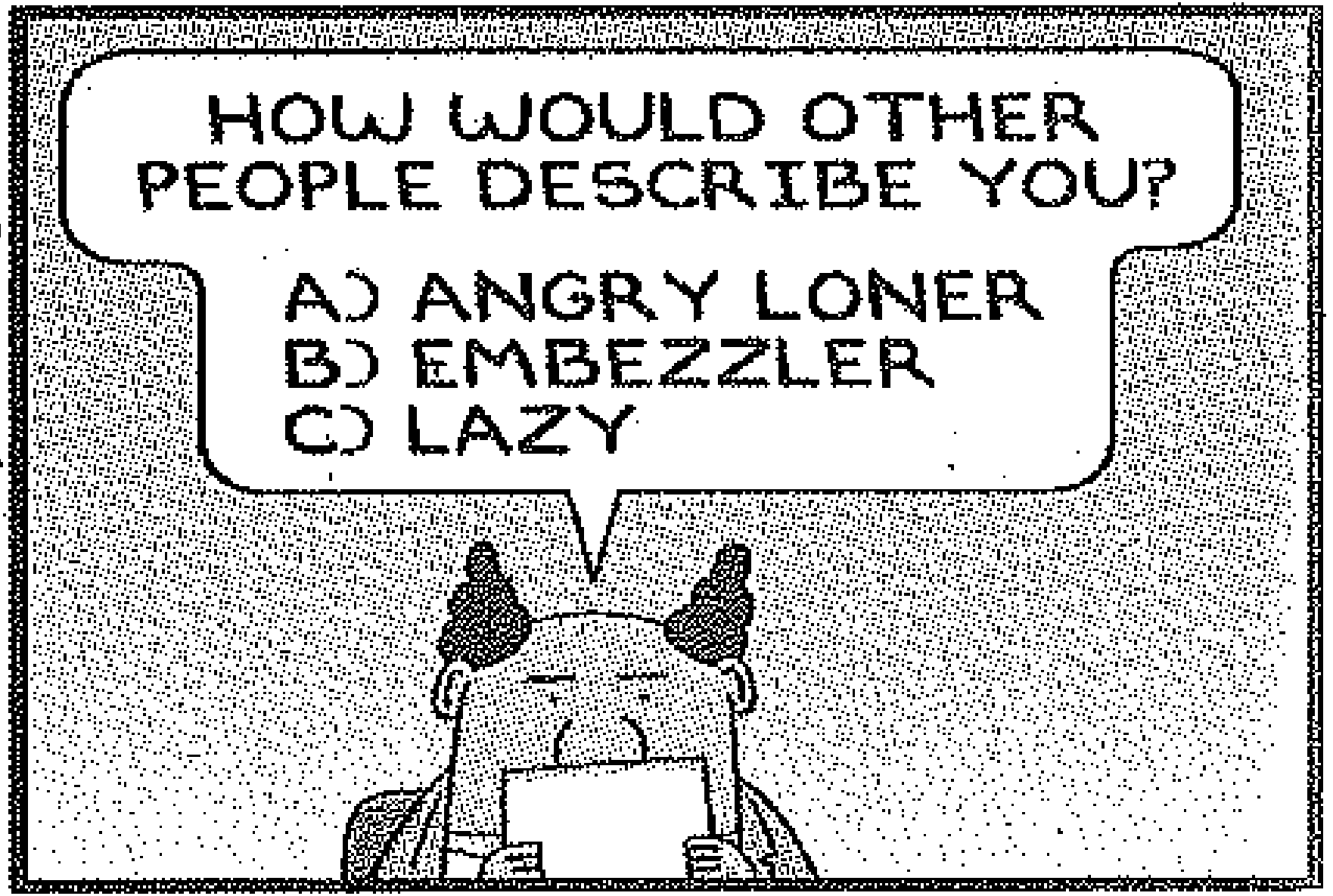
What can we do about faking?

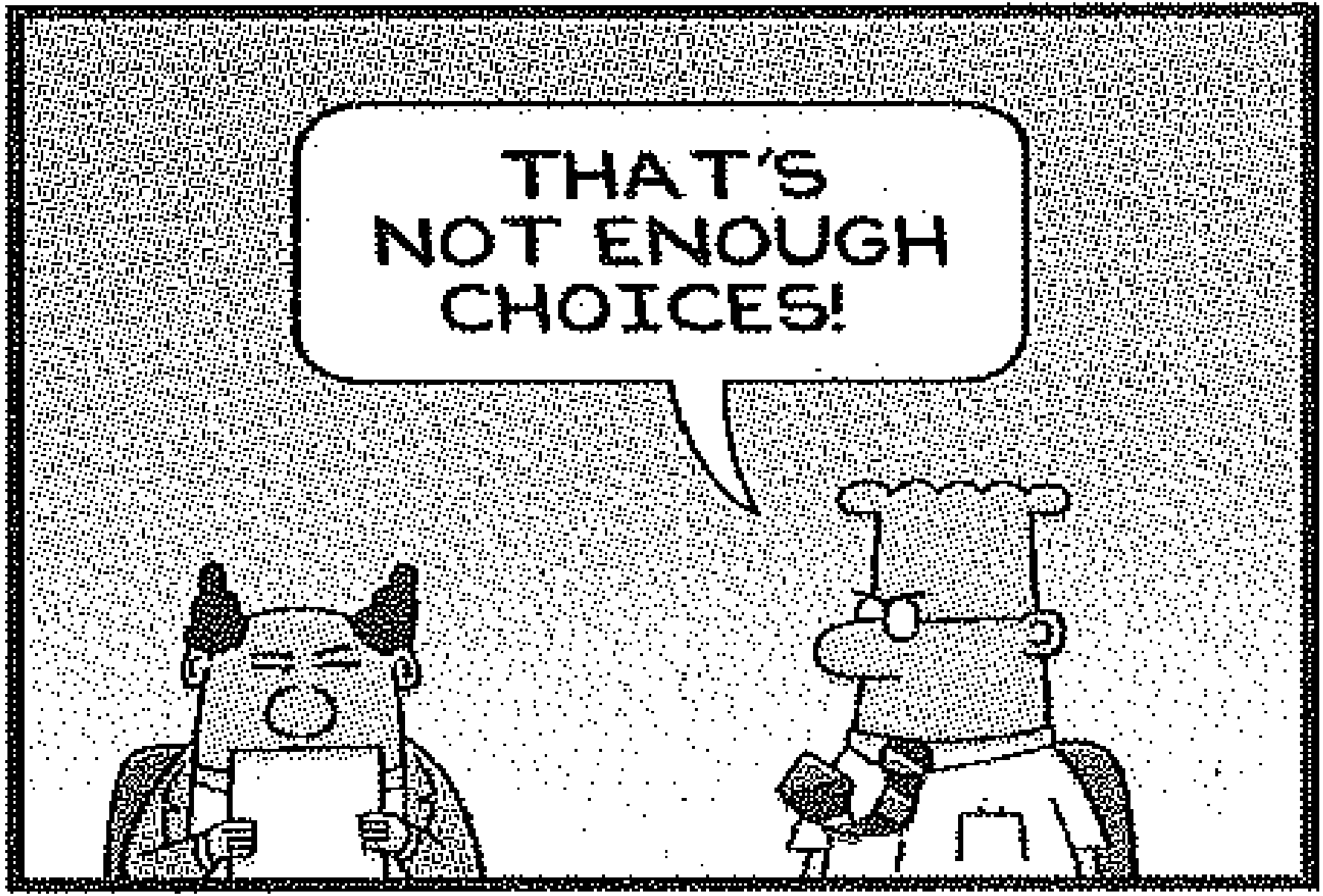
(Some) Prevention methods

- Warnings (pre-test or mid-test)
  - Pre-test warnings are claimed to reduce faking by 30-50% per applicant (Fan et al., 2012; Landers et al., 2011); others claim they are ineffective (Feeney et al., 2023)
  - Mid-test warnings elicit fear, guilt and anger (Li et al., 2022)
  - Warnings target the "faint hearted" thus putting them at a further disadvantage
- Item ambiguity
  - Obvious items are most susceptible to faking
  - Ambiguous items are not immune, eliciting faking in different directions (Kuncel & Borneman, 2007)
- Speeded tests
  - Reduce score inflation but only in low cognitive ability individuals (Komar et al., 2010)
  - Personality tests must not increase adverse impact!
- Forced choice and other ipsative response formats


WE'LL BE USING THE DOGBERT PERSONALITY PREDICTOR INDEX TO JUDGE YOUR CAREER POTENTIAL.

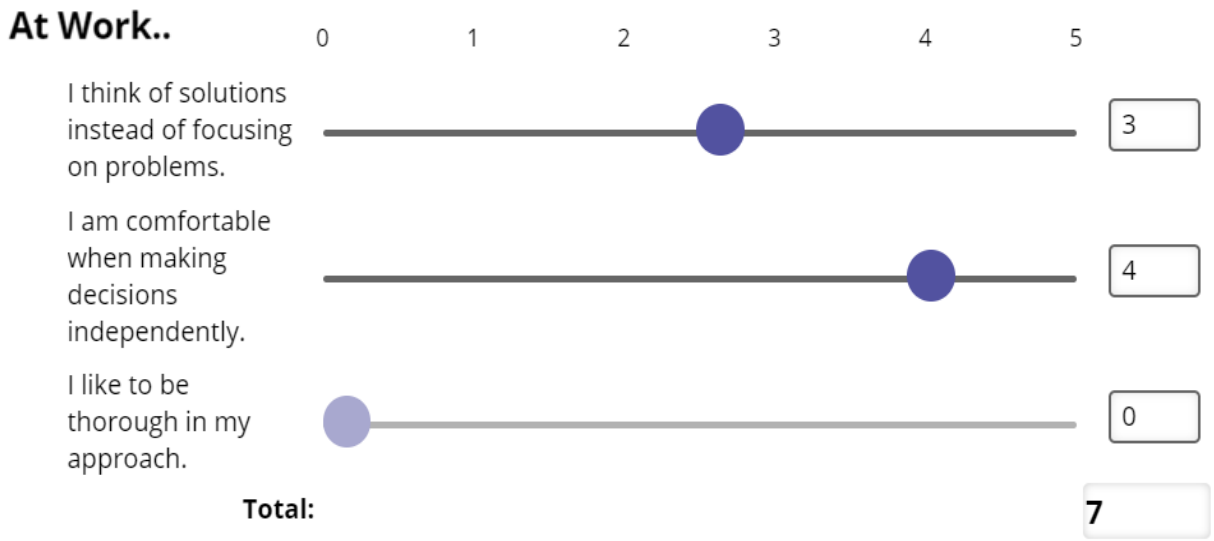

HERE'S A SAMPLE QUESTION. . .


HOW WOULD OTHER PEOPLE DESCRIBE YOU?
A) ANGRY LONER
B) EMBEZZLER
C) LAZY


THAT'S NOT ENOUGH CHOICES!


SAYS THE ANGRY LONER.

'Ipsative' response formats



- Impossible to endorse all desirable alternatives
- Facilitate differentiation and "slow" thinking (Kahneman, 2011)
- Popular since proper scaling methods have become available, e.g. Thurstonian models family (Brown & Maydeu-Olivares, 2011; 2013; 2018; Brown; 2016a; Brown; 2016b)
  - Normative trait scores can be obtained from ipsative data

Does Forced Choice (FC) reduce faking effects?

- Meta analyses show that the FC format reduces **effect sizes** of faking (Cao & Drasgow, 2019 ; Martinez and Salgado, 2021)
  - Many earlier controversies about this are test specific
  - FC is more effective in reducing faking effects when all choice alternatives appear equally desirable – "matched" blocks (Cao & Drasgow, 2019; Wetzel, Frick & Brown, 2020)
- **Construct validity** (e.g. factor structure) in medium and high stakes are better for FC (Brown, Inceoglu & Lin, 2017; Brown, 2023)
- **Operational validities** under faking conditions are higher for FC (e.g. Speer et al., 2023)

- These studies examined the effects (outcomes) of faking; they did not directly examine if faking was **prevented**

Does FC prevent faking?

A qualitative study by Fuechtenhans and Brown (2023) found that **matched desirable** and **matched undesirable** items act differently…

Prevent faking

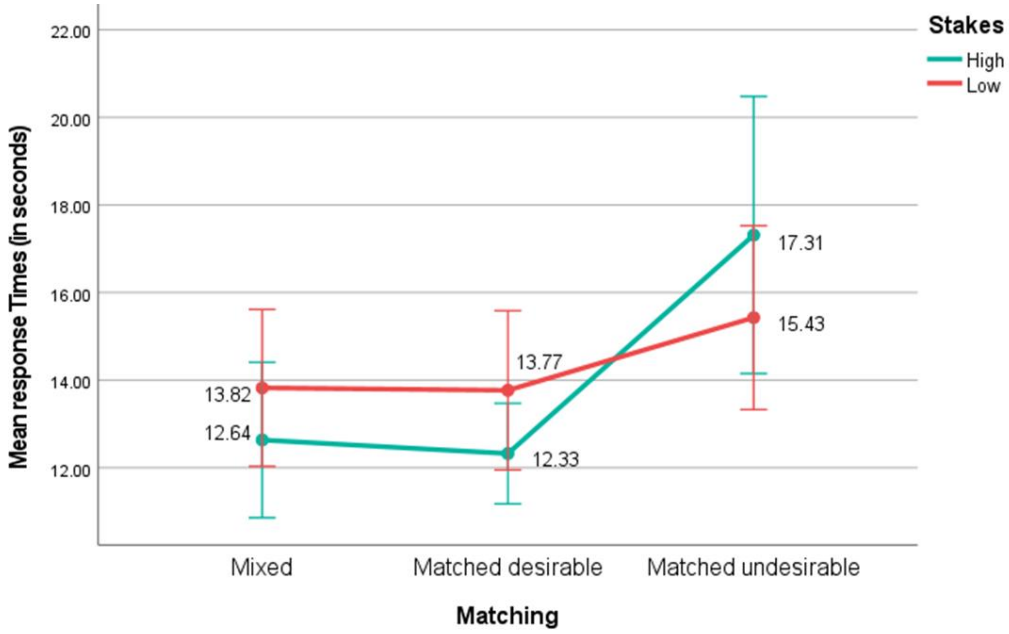
- When facing two **equally desirable** items, the respondent will fall back on true response (Gordon, 1951).
  - *"I, again went with an honest answer here, because there was nothing much to lose, it's all positive."*
  - *"…It was easy for me to answer because it's all good. So, I always choose the most truthful ones…"*

Facilitate faking

- Direct comparison of items facilitate acute differentiation of their desirability levels (Feldman & Corah, 1960).
- This was the case for **equally undesirable** items
  - *"…it was hard to choose which option was really me and tended to go with the one that my employer would be more likely to want."*
  - *"… I think the all-negative block is probably the hardest one. And one that you want to just make sure you pick the least negative…"*

Response latency as faking marker

- Response latencies have a long history of being markers of response manipulations, including faking (Holtgraves, 2004)
- Fuechtenhans (2023) conducted an experimental study, counterbalancing the order of stakes (low; high) and order of block matching (mixed, MD, MU), and found:
  - a significant **main** effect of matching (MU blocks took longer)
  - an interaction between matching and stakes
    - the difference between MU and other blocks was amplified in high stakes
- Matched undesirable items really facilitate more elaboration on response in high stakes!



Mixed keying vs. desirability matching?

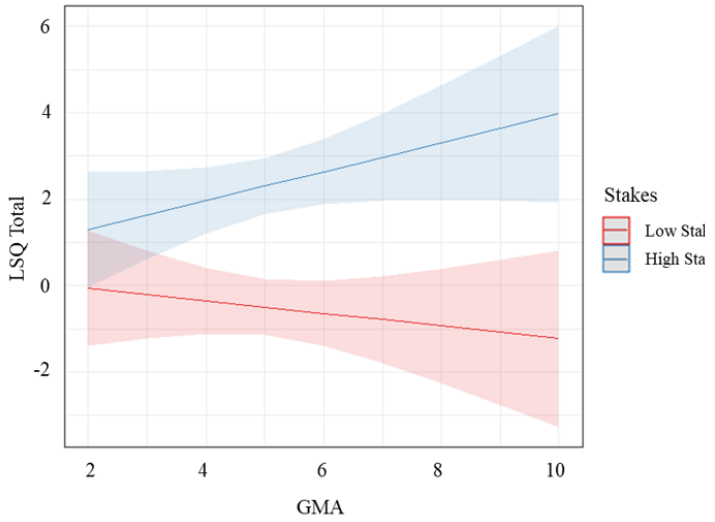
- To recover normative scores from ipsative data, factor loadings of items within comparisons, and within traits, must differ (Brown, 2016)
  - The **quickest** way to achieve that is to have some positively and some negatively keyed items
- Some argue that desirability matching makes it difficult to recover normative scores from ipsative data (e.g. Bürkner et al, 2019)
  - The condition of mixed keying is sufficient but not necessary! (Brown, 2016)
  - Li et al. (2024) show empirically that it is possible to strike a balance between social desirability matching and mixed keying, such that FC measures can have adequate psychometric properties and faking resistance.

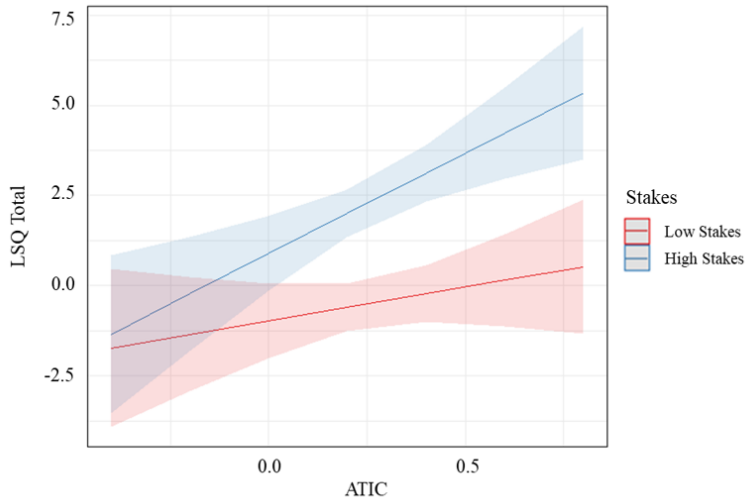
Ipsative questionnaires: The impact

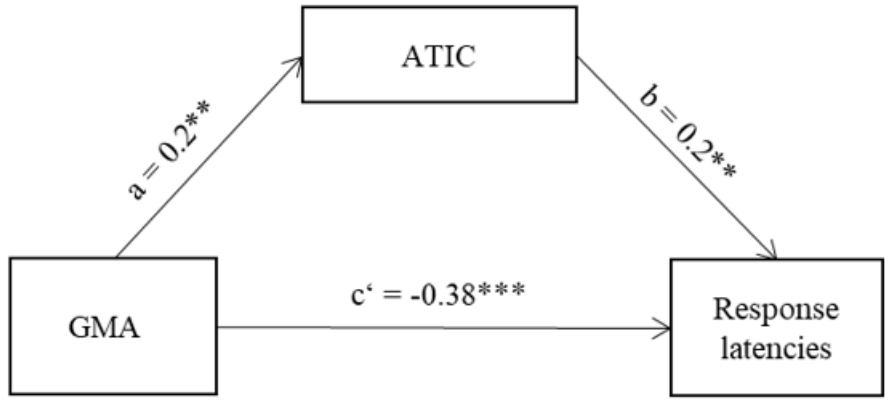
- Item Response Theory models such as MUPP (Stark et al., 2005) or TIRT (Brown & Maydeu-Olivares, 2011) have enabled the development of new generation of ipsative questionnaires enabling interpersonal comparisons
- The Thurstonian scaling approach has become the method of choice for development and scoring of high-stakes personality assessments internationally
  - Over 1,000,000 individual psychological assessments incorporating this methodology take place every year since 2011
  - in at least 39 different languages
  - across more than 40 different countries

Cognitive ability and faking ipsative questionnaires

- Research shows that cognitive ability plays a role in successfully faking FC measures (e.g. Vasilopoulos et al., 2006)
- Fuechtenhans (2024) investigated the role of abilities in faking a **graded-preference questionnaire** in an experimental within-subjects design with N=200
  - General Mental Ability (GMA); Ability to Identify Criteria (ATIC)







Bridging the gap between detection and prevention methods

What can we do about faking?

Detection & correction of faking for ipsative questionnaires

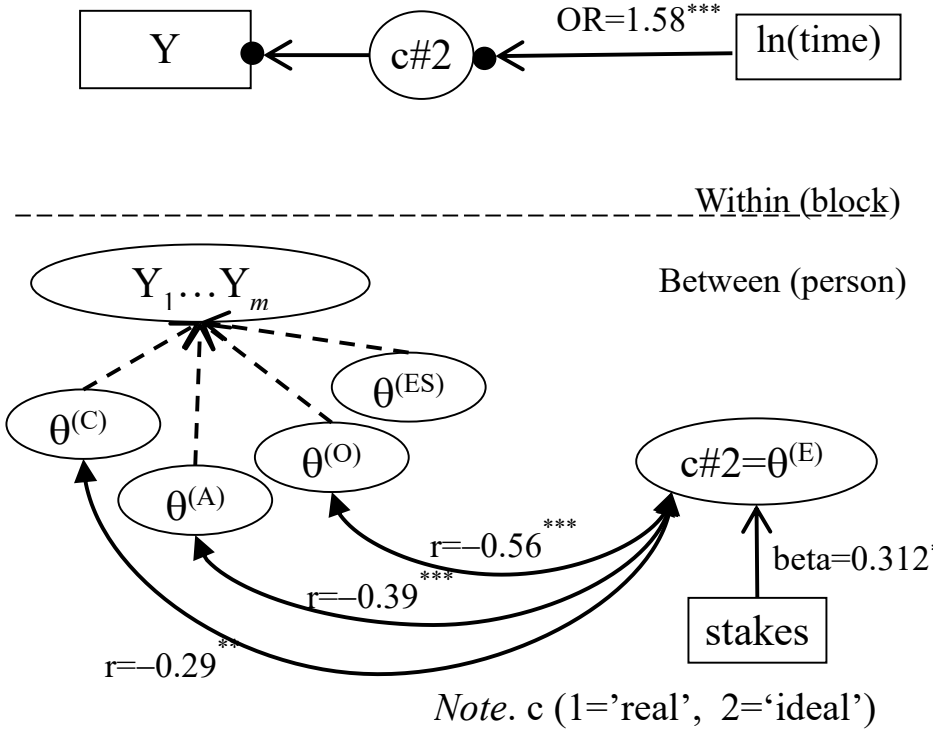
- We tend to detect & correct faking in normative questionnaires, and prevent it with ipsative
  - What about bridging the gap between correction and prevention methods?
- We need methodology for measuring faking on ipsative questionnaires
  - To evaluate the effectiveness of matching (and potentially correcting for the lack thereof)
- Frick (2022) described the "faking mixture" model allowing to measure "fakeability" of blocks, given that the item parameters in the honest condition are known (from lab data)
- Brown (2024) described and tested the "intermittent faking" model for ipsative questionnaires (F-GOMi) allowing to estimate all item and person parameters directly from high stakes (operational) data

Application of GOM analysis to ipsative high-stakes assessments

- In collaboration with John Eric Humphries & Katherine Kwok (Yale)

- An experimental study was embedded in the screening of software engineers by a US recruiter
- **Measure**: For a limited period, a purpose-built FC measure of Work Styles was administered, after programming tasks and biographic questions
- **Sample**: N=2,039 applicants (1,028 randomly assigned to HIGH stakes and 1,011 to LOW stakes)
  - To lower the stakes, applicants were told that completing the measure is voluntary and reassured that the scores will not influence selection decisions
- **Process data**: Response time; Number of clicks for each block

F-GOMi analysis: Results



- Class prevalence across both stakes
  - 75.3% of all responses classified as likely 'real'
  - 24.7% of all responses classified as likely 'ideal'
- WITHIN (item) level
  - **Longer** response times are associated with 'ideal' choices
- BETWEEN (person) level
  - HIGH stakes significantly **increase** propensity to report 'ideal' choices
  - Propensity to report 'ideal' choices correlates **negatively** with C, A and O

Conclusions

- Rating scales are open to faking, and we can attempt to control for it statistically (but beware consistent faking models!)
- Ipsative questionnaires cannot stop faking, but they can prevent it if constructed with understanding of the faking process
  - Make people want to tell the truth
  - Matched desirable items are the best; ambiguous items can also work
  - Matched undesirable items facilitate faking
- Cognitive ability plays a role in faking ipsative questionnaires
  - It is important to discourage faking, not to create adverse impact!
- Realistic models of faking such as the "intermittent faking" models can detect & control faking
  - Applying it to ipsative questionnaires can inform successful prevention

Future research directions

- Ipsative response formats
  - Ideal for research on General Factor of Personality (GFP). TIRT modelling retains substantive GFP!
- Further understanding of factors influencing faking
  - Person attributes (e.g. motivation, self-deception)
  - Item properties (e.g. valence, ambiguity and importance to criteria)
- Use process data to help understand faking
  - Response time, number of clicks, etc.
- Search for viable alternatives to self-report measures

Thank you!

Scan me!

Email: anna@annabrown.name

Visit my website

http://annabrown.name

“Psychometrics in Exercises using R and RStudio”

NEW textbook and data resource by Anna Brown

FREE access available from:
http://annabrown.name/book_psychometricsr/