Online Supplementary Material

The supplementary material comprises the statements used by Kurzban et al. (2001), a section on technical issues for the model of the WSW paradigm, and tables reporting the values of the model parameters.

Kurzban et al.'s (2001) Statements

1. You guys started the whole thing. That was the most flagrant foul I've ever seen. Your guy should have been ejected for that.
2. That's bullshit. You have to play the whistle. No whistle, no foul.
3. Hey, come on. He basically threw his elbow right into his face. You don't pull stuff like that when you're in our house.
4. Hey, you were the ones that started the fight. This whole thing wouldn't have happened if you could control yourselves.
5. You nail our guy in the face and expect to just get away with it? No way.
6. No one nailed anyone. It was a clean play. You guys get all bent out of shape, and now we're both screwed. Thanks for blowing our whole season.
7. Give me a break. We didn't blow your season. You did. Let's be serious.
8. Look, the point is you got out of control, went nuts, and got us on probation too. It's ridiculous.
9. If you just played like civilized people, the whole thing would never have happened.
10. You have to be kidding me. At least we don't play like you. You play like you're in high school.
11. And you play like you're in a zoo. Where you should be anyway.

12. Look, you’re just a bunch of wimps. You were sore because we beat you.

13. Yeah. After you took out our best guy by punching his lights out.

14. You guys sure complain a lot. You should do more playing and less whining.

15. Just shut up, man. You know the whole thing was your fault. You’re a bunch of animals.

16. Hey, yo, you better watch it. You need to cool off, friend.

17. Don’t be talking to us that way. You’re asking for some serious trouble.

18. Yeah? And who’s going to be making that trouble? You and the rest of the ladies here?

19. The only reason you’re not flat on your back right now is I don’t want to get kicked out before the game even starts.

20. Oh, now we’re really scared. What’s the matter with you people? You just can’t take it.

21. That’s about enough. Don’t you guys have something to do? In fact, why don’t you get back on the bus.

22. We’ll be leaving once we’ve gotten through with you. We’ll see who’s talking big after the game.

23. Well, if you guys won’t leave. We will. Come on, we’re out of here.

24. Fine. You guys take off. We’ll see you on the court.

Technical Issues

Correction Factors for Crossed Categories

When two categorizations are crossed, there is an additional issue with the error-difference measure. When guessing a speaker completely at random (i.e., where statement content does not bias one's guesses toward one or the other category, and each candidate speaker is as likely to be guessed as each other), the likelihood of a within-category error is smaller than the likelihood of a between-categories error. For example, if there are four White and four Black men, a Black man's statement can be assigned to three other Black men in a within-race error, but to four White men in a between-races error, implying that within-race errors are less likely by a factor of 3/4 than between-races errors. This is well known (Taylor et al., 1978) and usually corrected for by multiplying the between-races error rate with this factor in computing the error-difference measure. However, when race groups are further subdivided in subgroups of size 2 by, say, gender, a within-race within-gender error is only half as likely as any of the other three kinds of errors (within-race between-genders, between-races within-gender, between-races between-genders) and thus, when comparing the frequencies of these four kinds of errors between each other, the three kinds of errors other than the within-race within-gender errors should be multiplied with 1/2.

Vescio et al. (2004) propose to use the first correction (factor 3/4) when the error-difference measure is to be computed separately for each dimension collapsing across the other dimension. Vescio et al. (2004) use the second correction (factor 1/2) when comparing between the four different kinds of errors that can be defined at the subgroup level. In the case of crossed categories, the appropriate correction for the error-difference measure for one of the two crossed dimensions of categorization (collapsing across the other) is, however, somewhere between the two possibilities, and to further compound the issue, its size depends on the amount of category memory for the different dimensions of categorization.

To see this consider the case that there is no category memory of any kind for a given statement. In guessing a speaker completely at random, within-race errors are then indeed less likely by a factor of 3/4 than between-races errors. Consider a statement on the other hand, for which there is category memory for gender, but no memory for the speaker's race. Category memory for gender eliminates all between-genders errors, eliminating the four persons with the false gender and leaving the four members of the right gender as candidate speakers. In random guessing of one of these, a within-race error is now only half as likely as a between-races error, suggesting the use of the factor 1/2 as correction factor.

Whatever correction is used in computing the error-difference for one of two crossed dimensions of categorization, its appropriateness will therefore differ as a function of how much category memory there is for the other dimension. For example, for statements for which there is category memory for the second categorization, the second correction (factor 1/2) should be used in computing the error-difference measure for the first categorization. In contrast, for statements for which there is no category memory for the second categorization, the first correction (factor 3/4) should be used. The appropriate overall correction thereby actually falls between these two extremes in most cases depending on the amount of category memory there is for each dimension. Vescio et al. used the first correction (factor 3/4) comparing conditions with single and crossed categorization. This is the appropriate correction for the condition with single categorization. But in the crossed condition, between-errors are weighted more strongly than appropriate (the appropriate correction being smaller, between the two extremes of 1/2 and 3/4), reducing the size of the error-difference measure beyond what is appropriate for the crossed condition.

This same issue also has the potential to compromise comparisons between strong and weak dimensions of categorization as in Kurzban et al.'s (2001) studies. For example, there will frequently be category memory for the strong dimension (e.g., race), suggesting the use of the second correction (factor 1/2) in computing the error-difference measure for the weak categorization (e.g., team), but there will frequently be no category memory for the weak dimension, suggesting the use of the first correction (factor 3/4) in computing the error-difference measure for the strong categorization. But the appropriate correction actually falls between these two extremes in most cases depending on the amount of category memory there is for each dimension. There is no easy way out of this dilemma short of using the model-based approach proposed by Klauer et al. (2003), which solves this problem because the amount of category memory of each kind is estimated and thereby accounted for.

The Multi-Level Extension of the Mathematical Model

For a multilevel extension of the model, each parameter is considered to be a function of relevance condition with participants and items effects added to it to account for systematic differences between participants and items. For example, considering item memory, participants may differ in the extent to which they memorize the items, and items may differ in memorability. In the multi-level extension, the item-memory parameter I is given as I_{ijk}, where i indexes participants, j indexes items, and k indexes conditions with

$$I_{ijk} = G(\mu_k + \alpha_i + \beta_j), \tag{1}$$

where μ_k is the kth condition effect, α_i is the ith participant effect, and β_j is the jth item effect on item memory. The function G establishes a link between these effects and item memory. Such a link is necessary because item memory is a probability, constrained to range between 0 and 1, while the additive effects are real valued (Pratte & Rouder, 2012). Following Klauer (2010) and Matzke et al. (in press), we used a probit link, that is the inverse cumulative distribution function of a standard normal.

Remember that the baseline model sets equal parameters I, c, and d across statement type, and item effects for items were therefore also set equal across statement type (i.e., entered per statement pair) for these parameters. Failing to do so would have re-introduced differences in these parameters as a function of statement type via the backdoor of item effects.

Further constraints are gained by imposing hierarchical structures on participant and item effects for all model parameters. Participant and item effects are assumed to follow multivariate normal distributions with zero mean and a variance-covariance matrix that is a free parameter to be estimated from the data. There is one such matrix for the participant effects, and a separate one for item effects. These structures pull large effects toward zero, mitigating the influence of outliers. The variance-covariance matrices allow one to account for and estimate correlations between parameters across items and participants.

Estimation of the multilevel model can in principle be done via maximum-likelihood estimation, but the computational complexity of maximum-likelihood estimation is prohibitive for multilevel models (Klauer, 2010). We therefore used a Bayesian approach that is computationally tractable and converges with the maximum-likelihood estimates for large data samples. The Bayesian approach requires one to specify so-called hyperprior distributions for the condition effects and variance-covariance matrices. Here, we worked with the uninformative hyperpriors proposed by Matzke et al. (in press) and Klauer (2010). The parameter estimates reported in the tables and figures give the estimates for the condition effects. Lee and Wagenmakers (in press) and Rouder and Lu (2005) provide introductions to Bayesian hierarchical modeling of this kind.

Tables of Model Parameters

Table 1

Parameter Estimates and Credible Intervals in Study 2

	Team relevance		Low relevance		Race relevance	
Parameter	PE	95% CI	PE	95% CI	PE	95% CI
d(race)	0.05	[0.01,0.12]	0.16	[0.09,0.23]	0.22	[0.15,0.31]
d(team | race)	0.18	[0.06,0.42]	0.05	[0.01,0.21]	0.04	[0.00,0.20]
d(team | not race)	0.11	[0.02,0.19]	0.01	[0.00,0.03]	0.00	[0.00,0.01]
I	0.53	[0.46,0.59]	0.70	[0.64,0.76]	0.59	[0.52,0.68]
c	0.01	[0.00,0.03]	0.05	[0.02,0.08]	0.02	[0.00,0.04]
a_1(Black)	0.50	[0.46,0.53]	0.48	[0.45,0.52]	0.45	[0.41,0.49]
a_2(Black)	0.47	[0.44,0.51]	0.47	[0.44,0.51]	0.45	[0.41,0.49]
a_1(Team 1)	0.49	[0.46,0.52]	0.48	[0.45,0.51]	0.53	[0.50,0.56]
a_2(Team 1)	0.51	[0.47,0.55]	0.54	[0.50,0.58]	0.49	[0.44,0.52]
b_1	0.27	[0.20,0.35]	0.18	[0.11,0.26]	0.22	[0.15,0.29]
b_2	0.23	[0.18,0.31]	0.13	[0.07,0.19]	0.13	[0.08,0.19]

Note. CI = credible interval; PE = parameter estimate. Team 1 is the team with red jackets. Parameters a_i and b_i refer to statements of Type i. Statements of the Type 1 ($i = 1$) are for high team, low, and high race relevance, in order, own-team aggrandizing, conservative, and claiming the absence of race-based differences.

Table 2

Parameter Estimates and Credible Intervals in Study 3

	Team relevance		Low relevance		Race relevance	
Parameter	PE	95% CI	PE	95% CI	PE	95% CI
d(race)	0.07	[0.03,0.15]	0.25	[0.15,0.34]	0.38	[0.28,0.48]
d(team | race)	0.12	[0.03,0.30]	0.10	[0.01,0.29]	0.11	[0.05,0.22]
d(team | not race)	0.16	[0.06,0.24]	0.01	[0.00,0.09]	0.04	[0.00,0.16]
I	0.50	[0.42,0.59]	0.55	[0.45,0.65]	0.53	[0.44,0.62]
c	0.03	[0.01,0.07]	0.05	[0.02,0.09]	0.03	[0.01,0.06]
a_1(Black)	0.47	[0.43,0.52]	0.48	[0.44,0.52]	0.48	[0.44,0.53]
a_2(Black)	0.48	[0.43,0.53]	0.46	[0.41,0.51]	0.49	[0.43,0.55]
a_1(Team 1)	0.51	[0.47,0.55]	0.50	[0.46,0.54]	0.48	[0.44,0.51]
a_2(Team 1)	0.49	[0.45,0.54]	0.53	[0.48,0.57]	0.52	[0.47,0.58]
b_1	0.19	[0.12,0.27]	0.13	[0.09,0.20]	0.13	[0.08,0.18]
b_2	0.14	[0.08,0.21]	0.07	[0.04,0.16]	0.10	[0.06,0.17]

Note. CI = credible interval; PE = parameter estimate. Team 1 is the team with green jackets. Parameters a_i and b_i refer to statements of Type i. Statements of Type 1 ($i = 1$) are for high team, low, and high race relevance, in order, own-team aggrandizing, satisfied, and claiming the absence of race-based differences.

Table 3

Parameter Estimates and Credible Intervals in Study 4

	Without Comparative Focus		With Comparative Focus	
Parameter	PE	95% CI	PE	95% CI
d(race)	.28	[.22,.37]	.17	[.10,.24]
d(prison | race)	.09	[.01,.33]	.24	[.05,.73]
d(prison | not race)	.06	[.01,.13]	.01	[.00,.05]
I	.84	[.79,.88]	.66	[.58,.73]
c	.07	[.04,.12]	.04	[.01,.08]
a_1(Black)	.50	[.46,.54]	.49	[.46,.53]
a_2(Black)	.53	[.47,.59]	.56	[.49,.61]
a_1(Prison 1)	.52	[.48,.56]	.48	[.44,.52]
a_2(Prison 1)	.55	[.50,.61]	.50	[.46,.55]
b_1	.05	[.02,.09]	.08	[.04,.14]
b_2	.07	[.04,.15]	.09	[.04,.16]

Note. CI = credible interval; PE = parameter estimate. Prison 1 is the Wakefield prison. Parameters a_i and b_i refer to statements of Type i. Statements of Type 1 ($i = 1$) express satisfaction with the (old) prison.

Table 4

Parameter Estimates and Credible Intervals in Study 5

	Team relevance		Low relevance		Gender relevance	
Parameter	PE	95% CI	PE	95% CI	PE	95% CI
d(gender)	0.14	[0.07,0.22]	0.49	[0.39,0.59]	0.66	[0.56,0.73]
d(team | gender)	0.16	[0.00,0.37]	0.02	[0.00,0.07]	0.12	[0.03,0.21]
d(team | not gender)	0.14	[0.07,0.26]	0.01	[0.00,0.04]	0.02	[0.00,0.10]
I	0.55	[0.48,0.62]	0.60	[0.53,0.67]	0.65	[0.59,0.73]
c	0.03	[0.01,0.06]	0.10	[0.05,0.14]	0.18	[0.13,0.23]
a_1(Female)	0.51	[0.45,0.56]	0.50	[0.45,0.55]	0.52	[0.46,0.57]
a_2(Female)	0.50	[0.45,0.55]	0.40	[0.32,0.47]	0.48	[0.40,0.56]
a_1(Team 1)	0.53	[0.49,0.58]	0.52	[0.48,0.56]	0.48	[0.43,0.52]
a_2(Team 1)	0.58	[0.53,0.63]	0.51	[0.46,0.59]	0.49	[0.41,0.55]
b_1	0.18	[0.10,0.26]	0.06	[0.03,0.12]	0.09	[0.05,0.18]
b_2	0.15	[0.10,0.23]	0.08	[0.05,0.14]	0.11	[0.05,0.19]

Note. CI = credible interval; PE = parameter estimate. Team 1 is the team with green jackets. Parameters a_i and b_i refer to statements of Type i. Statements of Type 1 ($i = 1$) are for high team, low, and high gender relevance, in order, own-team aggrandizing, satisfied, and conservative.