

## Online Supplemental Materials

### Representativeness of the Sample in Studies 1 and 2

The 597 observations of Study 1 ( $M_{\text{age}} = 44$ ,  $SD = 11$ ; 67% male) were comparable in their demographics to the total group of 4,149 race registrants ( $M_{\text{age}} = 44$ ,  $SD = 13$ ; 63% male); the 369 observations of Study 2 ( $M_{\text{age}} = 46$ ,  $SD = 11$ ; 67% male) were comparable to the total group of 4,716 race registrants ( $M_{\text{age}} = 45$ ,  $SD = 13$ ; 63% male).

### Covariate Measures Used in Studies 1 and 2

As covariate measures, we included the runners' *gender* ( $0 = \text{male}$ ,  $1 = \text{female}$ ), *age* (*year of birth*), and *running expertise*. The latter consisted of 7 items: i. number of years of competitive running ("For how many years do you run competitions?"), ii. number of prior half marathons ("In how many halfmarathon races have you participated so far? (Rough estimation; without your upcoming participation in the [name of race])"), iii. number of prior marathons ("In how many marathon races have you participated so far? (Rough estimation; without your upcoming participation in the [name of race])"), iv. average training km per week ("How many kilometers per week did you run on average as training? (km/week)"), v. personal marathon best time<sup>1</sup> ("What is your personal record in a marathon competition? (in hh.mm, e.g., 4.30)"), vi. personal half marathon best time ("What is your personal record in a half marathon competition? (in hh.mm, e.g., 2.30)"), and vii. estimated 10k race-like pace ("How fast do you currently run ten kilometers (10km, competition-like race pace, flat)? (in mm.ss, e.g., 60.00)"). All items were

---

<sup>1</sup> A programming error in Study 1 prevented the first 6% of participants from entering their marathon and half marathon best times, meaning that for those 6%, two of the seven items comprising the covariate composite score were missing.

rank transformed using percentiles before an individual's composite running-expertise score was computed ( $\alpha_{\text{Study}1/2} = .72/.80$ ;  $\text{Range}_{\text{Study}1/2} = 5.81/6.32$  to  $93.12/98.43$ ). If a runner had no prior marathon or half marathon best time, the composite score for those runners was computed without these variables.

### **Additional Measures Elicited in Studies 1 and 2**

*Motivational measures.* How likely do you think it is that you will attain your predicted time? (0 = not at all likely, 6 = very likely); How important is it to you that you will attain your predicted time? (0 = not at all important, 6 = very important); How disappointed would you feel if you did not attain your predicted time? (0 = not at all disappointed, 6 = very disappointed); How hard will you try to attain your predicted time? (0 = not at all hard, 6 = very hard). (Adapted from Oettingen & Mayer, 2002; Oettingen, Pak, & Schnetter, 2001)

*Emotion measures.* How do you feel today if you think of the upcoming [name of the event] race? (Please indicate how strongly you feel the following emotions from 0 = not at all, 6 = very strongly. If I think of the upcoming race today, I feel:) 10 emotion items: optimistic, confident, joyful, hopeful, worried, anxious, uncomfortable, hopeless, calm, tense). (Derived from Baumgartner, Pieters, & Bagozzi 2008; Leone, Perugini, & Bagozzi 2005; Pekrun 2006)

*Cognitive measures.* What/who did you use as reference to make this judgment about the average [name of the event] runner? (0 = not at all, 6 = very much; 4 items: myself, my idea about typical runners, a specific runner that I know, other references [please specify]); How similar do you perceive yourself compared to the average [name of race] runner (you answered the question for)? (0 = very dissimilar, 6 = very similar). (Adapted from Engeler & Raghubir, 2018)

*Running-specific measures.* What applies to you most? (4 choices: Non-ambitious hobby runner, Ambitious hobby runner, Amateur / Semi-professional runner, Professional runner);

Which main goal do you pursue for your race? (2 choices: To finish the race at all; To finish the race in a certain time).

*Additional demographic variable.* In what country do you live?

In Study 1 only, we additionally included questions in a post-survey that was sent out by the organizer after the race. These were: How satisfied are you with your race performance? (0 = not at all satisfied, 6 = very satisfied); How hard did you try to achieve your time goal? (0 = not at all hard, 6 = very hard); How much did you enjoy your race? (0 = did not at all enjoy it, 6 = enjoyed it very much); How did the race go? (0 = not at all ideal, 6 = everything was ideal). The remainder of that post-race questionnaire included questions of interest to the race organizer, such as about event satisfaction, expenditure, and travel behavior.

### Meta-Analysis Without Covariates

Table S1 below shows the results of the meta-analysis without covariates. The substantive conclusions are the same as those from the meta-analysis reported in Table 2 in the paper.

Table S1

#### *Miscalibration of Predictions About One's Own Performance Across Studies.*

Type of (mis)placement	Study	Mean	Lower 95% CI	Upper 95% CI	SD	<i>t</i>	<i>p</i>	<i>d</i>	Average <i>d</i>	<i>Z</i> Meta-Analysis	<i>p</i> Meta-Analysis
<i>Erroneous WTA</i>	Study 1	2.69	0.10	5.28	11.26	2.07	.042	0.24			
	Study 2	5.56	1.31	9.81	15.11	2.63	.011	0.37			
	M&S (pretask)	0.48	-0.78	1.74	3.78	0.78	.442	0.13			
	M&S (posttask)	0.03	-0.34	0.41	1.29	0.18	.859	0.03			
	<b>Overall</b>								<b>0.19</b>	<b>2.29</b>	<b>.022</b>
<i>Correct WTA</i>	Study 1	-4.42	-6.04	-2.80	11.64	-5.39	< .001	-0.38			
	Study 2	-3.17	-5.30	-1.04	11.98	-2.95	.004	-0.26			
	M&S (pretask)	-1.39	-2.67	-0.12	4.14	-2.20	.033	-0.34			
	M&S (posttask)	-0.79	-1.15	-0.42	1.64	-4.28	< .001	-0.48			
	<b>Overall</b>								<b>-0.36</b>	<b>3.43</b>	<b>&lt; .001</b>
<i>Erroneous BTA</i>	Study 1	-12.81	-16.31	-9.31	12.18	-7.36	< .001	-1.05			
	Study 2	-16.97	-23.11	-10.83	13.12	-5.79	< .001	-1.29			
	M&S (pretask)	-3.43	-4.42	-2.45	3.30	-7.05	< .001	-1.04			
	M&S (posttask)	-0.95	-1.60	-0.30	1.98	-2.96	.005	-0.48			
	<b>Overall</b>								<b>-0.97</b>	<b>3.46</b>	<b>&lt; .001</b>
<i>Correct BTA</i>	Study 1	-2.62	-3.94	-1.30	9.56	-3.92	< .001	-0.27			
	Study 2	-3.21	-5.18	-1.24	10.22	-3.23	.002	-0.31			
	M&S (pretask)	0.36	-0.52	1.25	3.49	0.82	.418	0.10			
	M&S (posttask)	-0.15	-0.43	0.13	1.21	-1.06	.291	-0.12			
	<b>Overall</b>								<b>-0.15</b>	<b>1.84</b>	<b>.066</b>

*Note.* Positive means indicate that participants performed *better* than predicted (*underestimation* of performance), and negative means indicate that they performed *worse* than predicted (*overestimation* of performance). *t* values pertain to one-sample *t* tests against 0. *Z*<sub>Meta-Analysis</sub> tests the mean *p* values (Rosenthal, 1978;  $Z = [.50 - \bar{p}] [\sqrt{12N}]$ , *N* is number of studies) across all four data points (Studies 1 and 2; M&S Pre- and Posttask). M&S denotes Moore & Small's (2007) Study 1 data. BTA = better than average; WTA = worse than average.

**Source of Misplacement: Means**

Table S2 shows the means of misestimation (self vs. other) across all cells.

Table S2

*Miscalibration of Predictions About One's Own and Others' Performance Across Studies.*

Type of (mis)placement	Study	N	Misestimation Performance Self			Misestimation Performance Other		
			Mean	Lower 95% CI	Upper 95% CI	Mean	Lower 95% CI	Upper 95% CI
<i>Erroneous WTA</i>	Study 1	75	1.60	-0.92	4.12	-16.63	-19.61	-13.65
	Study 2	51	3.58	0.28	6.88	-15.07	-18.41	-11.74
	M&S (pretask)	37	0.48	-0.78	1.74	-1.26	-2.53	0.01
	M&S (posttask)	48	0.03	-0.34	0.41	-1.84	-2.30	-1.37
<i>Correct WTA</i>	Study 1	201	-5.05	-6.64	-3.45	-12.40	-14.29	-10.52
	Study 2	124	-4.65	-6.87	-2.42	-9.71	-11.95	-7.46
	M&S (pretask)	43	-1.39	-2.67	-0.12	-1.32	-2.56	-0.09
	M&S (posttask)	80	-0.79	-1.15	-0.42	-1.35	-1.83	-0.86
<i>Erroneous BTA</i>	Study 1	49	-12.74	-15.81	-9.67	5.67	2.04	9.30
	Study 2	20	-14.62	-19.81	-9.43	0.16	-5.07	5.40
	M&S (pretask)	46	-3.43	-4.42	-2.45	-1.64	-2.73	-0.55
	M&S (posttask)	38	-0.95	-1.60	-0.30	0.85	0.21	1.49
<i>Correct BTA</i>	Study 1	205	-1.61	-3.23	0.01	-4.54	-6.45	-2.63
	Study 2	106	-1.20	-3.71	1.30	-4.29	-6.81	-1.77
	M&S (pretask)	62	0.36	-0.52	1.25	0.03	-0.85	0.90
	M&S (posttask)	73	-0.15	-0.43	0.13	-0.49	-0.82	-0.16

*Note.* Positive means indicate self/other performed *better* than predicted (*underestimation* of performance), and negative means indicate self/other performed *worse* than predicted (*overestimation* of performance). M&S denotes Moore & Small's (2007) Study 1 data.

BTA = better than average; WTA = worse than average.

## Scatterplots of Raw Data (All Studies)

## Study 1

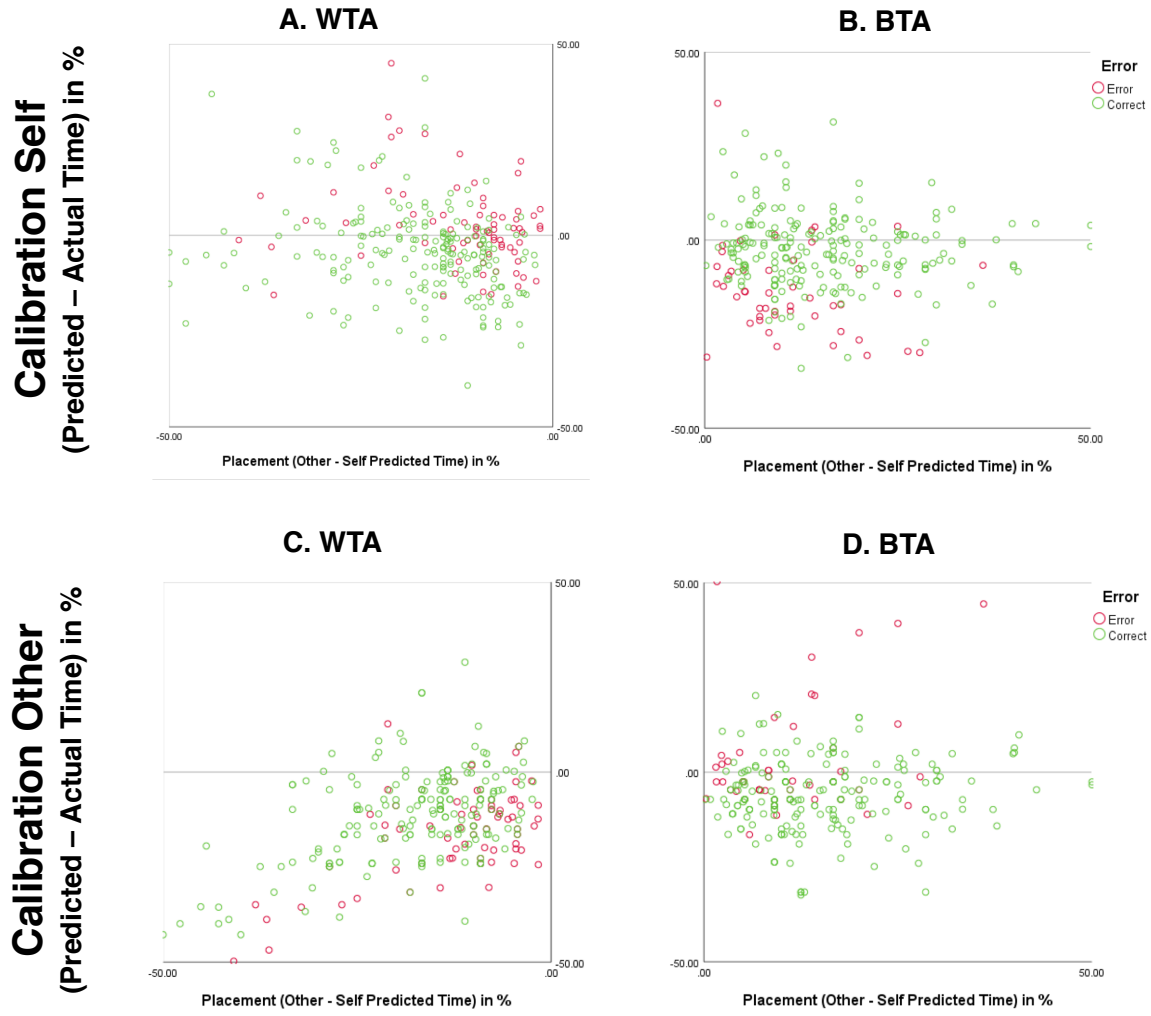


Figure S1. Scatterplot of Raw Data Points for Misestimation of Own and Others' Absolute Performance (Study 1). Y-axis denotes the degree of misestimation: Positive values mean finishing *faster* than predicted (*underestimation* of performance), and negative values mean finishing *slower* than predicted (*overestimation* of performance). X-axis denotes the degree of misplacement: Positive values (Panel B. and D.) mean the self is predicted to finish *faster* than the other, and negative values (Panel A. and C.) mean the self is predicted to finish *slower* than

the other. For ease of presentation, five values are outside of the displayed range for Calibration Self and eight values for Calibration Other.

The raw data scatterplots in Figure S1 visualize the placement belief by belief correctness interaction discussed in the paper across the different levels of self-other differences. Among those who rated themselves BTA (Panel B), those with correct BTA beliefs were more accurate about their own performance—shown by the green dots hovering along the zero line. In contrast, those with erroneous BTA beliefs tended to overestimate their own performance—shown by the majority (87.8%) of the red dots being below the zero line. Their erroneous beliefs that they are BTA when they were not stem primarily from their overestimating their own performance; they erred less about others' performance, which, in Study 1, they tended to underestimate—shown by the majority (57.1%) of red dots in Panel D being above the zero line (with many of the red dots above the zero line having a larger distance from it than the red dots below).

Among those who rated themselves WTA (Panel A), the pattern was reversed. It was those with correct WTA beliefs who tended to overestimate their own performance—shown by the majority (70.1%) of green dots being below the zero line. In contrast, those with erroneous WTA beliefs tended to be more accurate about their own performance; if anything, they tended to underestimate it—shown by the majority (57.3%) of the red dots being above the zero line. Their erroneous beliefs that they are WTA when they were not was primarily because they tended to overestimate others' performance—shown by the majority (94.7%) of red dots in Panel C being below the zero line.

## Study 2

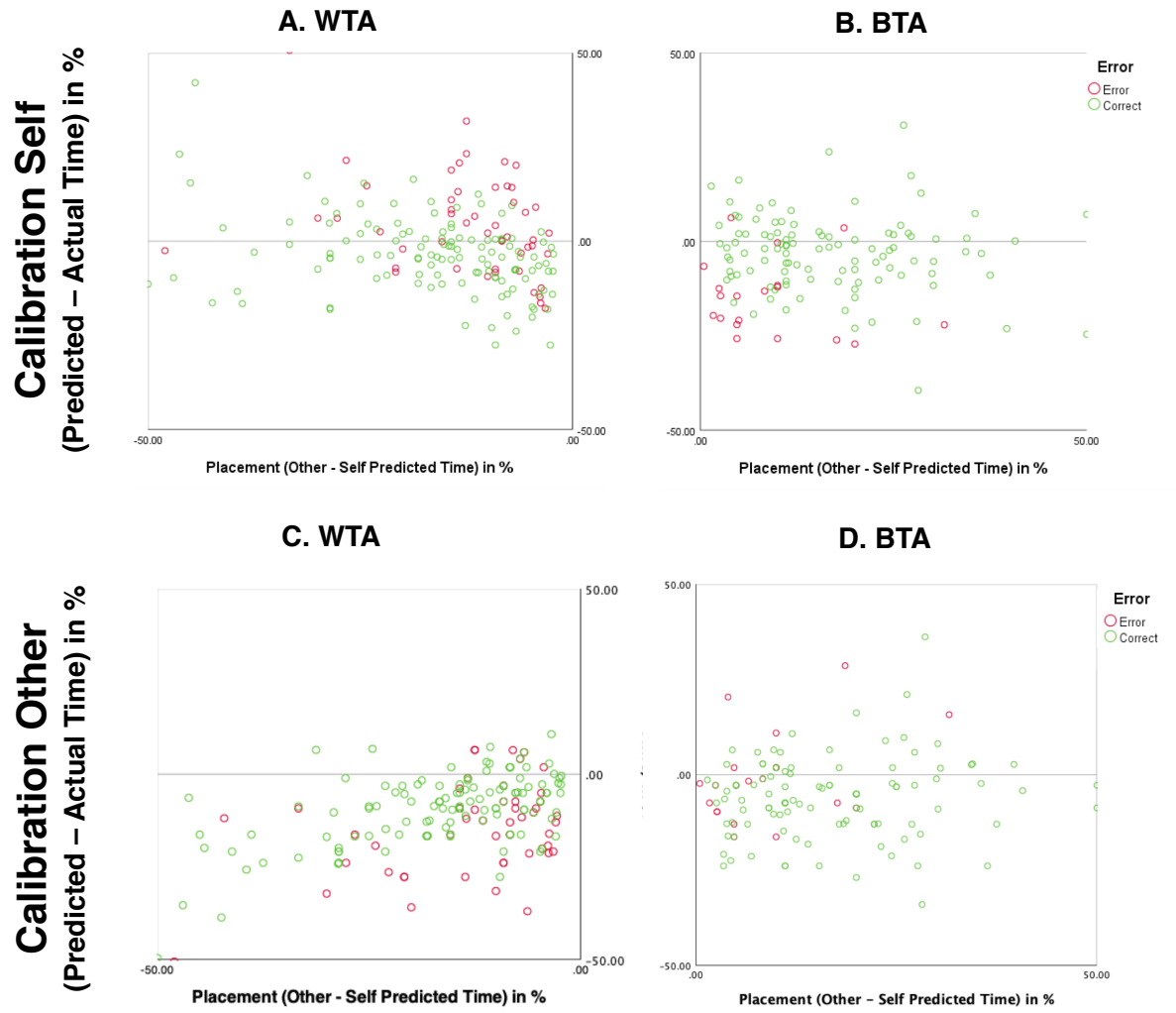
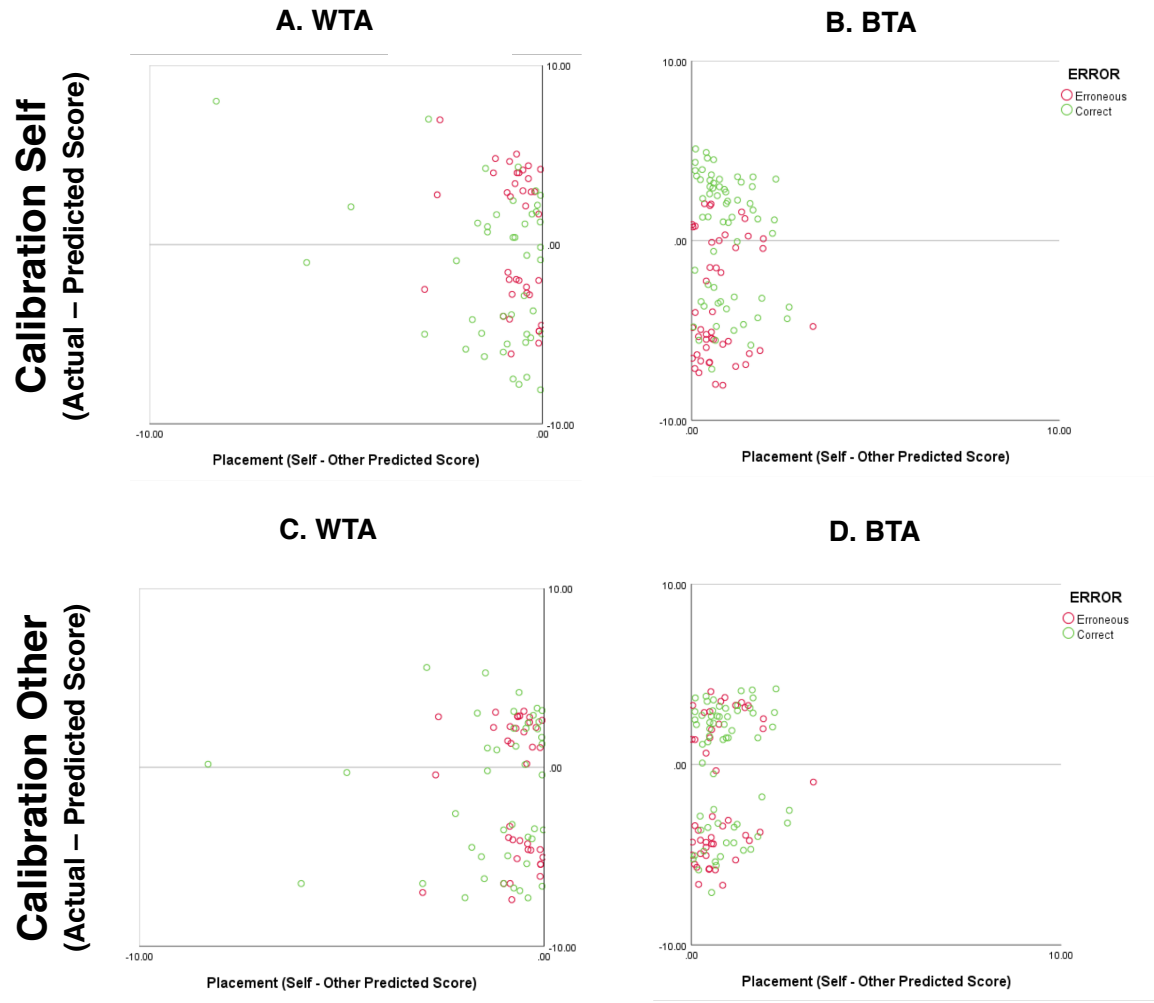


Figure S2. Scatterplot of Raw Data Points for Misestimation of Own and Others' Absolute Performance (Study 2). Y-axis denotes the degree of misestimation: Positive values mean finishing *faster* than predicted (*underestimation* of performance), and negative values mean finishing *slower* than predicted (*overestimation* of performance). X-axis denotes the degree of misplacement: Positive values (Panel B. and D.) mean the self is predicted to finish *faster* than the other, and negative values (Panel A. and C.) mean the self is predicted to finish *slower* than the other. For ease of presentation, five values are outside of the displayed range for Calibration Self and three values for Calibration Other.



The raw data scatterplots in Figure S2 visualize the placement belief by belief correctness interaction discussed in the paper across the different levels of self-other differences. Among those who rated themselves BTA (Panel B), those with correct BTA beliefs were more accurate about their own performance—shown by the green dots hovering along the zero line. In contrast, those with erroneous BTA beliefs tended to overestimate their own performance—shown by the majority (90.0%) of the red dots being below the zero line. Their erroneous beliefs that they are BTA when they were not stem primarily from their overestimating their own performance; they erred less about others' performance, which, in Study 2, they tended to be accurate about—shown by the red dots in Panel D hovering around the zero line.

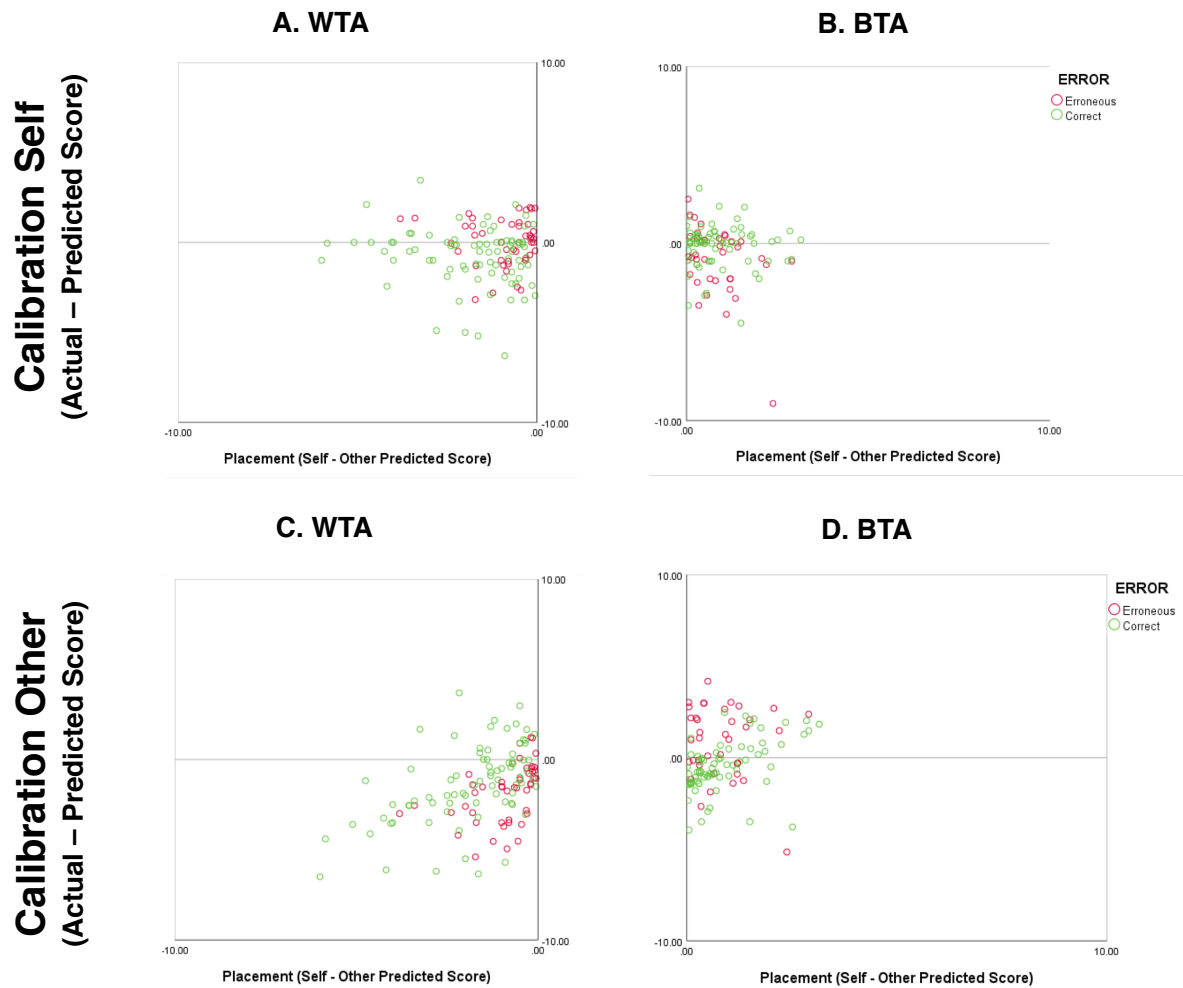
Among those who rated themselves WTA (Panel A), the pattern was reversed. It was those with correct WTA beliefs who tended to overestimate their own performance—shown by the majority (66.1%) of green dots being below the zero line. In contrast, those with erroneous WTA beliefs tended to be more accurate about their own performance; if anything, they tended to underestimate it—shown by the majority (62.7%) of the red dots being above the zero line. Their erroneous beliefs that they are WTA when they were not was primarily because they tended to overestimate others' performance—shown by the majority (88.2%) of red dots in Panel C being below the zero line.

**Moore & Small (2007), Pretask Predictions**

*Figure S3.* Scatterplot of Raw Data Points for Misestimation of Own and Others' Absolute Performance (Moore & Small, 2007, Pretask predictions). Y-axis denotes the degree of misestimation: Positive values mean performing *better* than predicted (*underestimation* of performance), and negative values mean performing *worse* than predicted (*overestimation* of performance). X-axis denotes the degree of misplacement: Positive values (Panel B. and D.) mean the self is predicted to perform *better* than the other, and negative values (Panel A. and C.) mean the self is predicted to perform *worse* than the other.

The raw data scatterplots in Figure S3 visualize the placement belief by belief correctness interaction discussed in the paper across the different levels of self-other differences. Among those who rated themselves BTA (Panel B), those with correct BTA beliefs were more accurate about their own performance—shown by the green dots hovering along the zero line. In contrast, those with erroneous BTA beliefs tended to overestimate their own performance—shown by the majority (73.9%) of the red dots being below the zero line. Their erroneous beliefs that they are BTA when they were not stem primarily from their overestimating their own performance; they erred less about others' performance, which they tended to overestimate as well—shown by a smaller majority (60.9%) of red dots in Panel D being above the zero line.

Among those who rated themselves WTA (Panel A), the pattern was reversed. It was those with correct WTA beliefs who tended to overestimate their own performance—shown by the majority (55.8%) of green dots being substantially below the zero line. In contrast, those with erroneous WTA beliefs tended to be more accurate about their own performance; if anything, they tended to underestimate it—shown by the majority (54.1%) of the red dots being above the zero line. Their erroneous beliefs that they are WTA when they were not was primarily because they tended to overestimate others' performance—shown by a large part (48.6%) of red dots in Panel C being substantially below the zero line.

**Moore & Small (2007), Posttask predictions**

*Figure S4.* Scatterplot of Raw Data Points for Misestimation of Own and Others' Absolute Performance (Moore & Small, 2007, Posttask predictions). Y-axis denotes the degree of misestimation: Positive values mean performing *better* than predicted (*underestimation* of performance), and negative values mean performing *worse* than predicted (*overestimation* of performance). X-axis denotes the degree of misplacement: Positive values (Panel B. and D.) mean the self is predicted to perform *better* than the other, and negative values (Panel A. and C.) mean the self is predicted to perform *worse* than the other.

The raw data scatterplots in Figure S4 visualize the placement belief by belief correctness interaction discussed in the paper across the different levels of self-other differences. Among those who rated themselves BTA (Panel B), those with correct BTA beliefs were more accurate about their own performance—shown by the green dots hovering along the zero line. In contrast, those with erroneous BTA beliefs tended to overestimate their own performance—shown by the majority (63.2%) of the red dots being below the zero line. Their erroneous beliefs that they are BTA when they were not stem primarily from their overestimating their own performance; the majority (63.2%) also tended to underestimate others' performance—shown by the red dots in Panel D being above the zero line.

Among those who rated themselves WTA (Panel A), the pattern was reversed. It was those with correct WTA beliefs who tended to overestimate their own performance—shown by the majority (66.3%) of green dots being substantially below the zero line. In contrast, those with erroneous WTA beliefs tended to be more accurate about their own performance; if anything, they tended to underestimate it—shown by a large part (50.0%) of the red dots being substantially above the zero line. Their erroneous beliefs that they are WTA when they were not was primarily because they tended to overestimate others' performance—shown by the majority (89.6%) of red dots in Panel C being substantially below the zero line.

### **Reexamination of Moore and Healy's (2008) Study**

We reexamined the data from Moore and Healy's (2008) study ( $N = 82$  U.S. students; 51% male) through a placement belief by belief correctness framework. As in Moore and Small (2007), participants were instructed to take a trivia quiz of 10 questions and estimated the probability of attaining each of the 11 possible scores (from 0 correct to 10 correct), both for

themselves and for another person, again prior to and after the task. The key dependent variable—calibration of participants’ predictions of their own absolute performance—and the placement belief by belief correctness classification of each participant were computed as in the reexamination of the Moore and Small (2007) data.

There were several differences in the study design compared to Moore and Small’s (2007, Study 1) trivia study (for which reason this study was not included in the meta-analysis). First, the questions had three (not two) difficulty levels: easy, medium, difficult. Second, participants were asked to play 18 rounds (not 1 round) of 10 questions, which provided 18 observations (not 1 observation) per participant and hence 1,476 observations for the entire sample. Therefore, predictions from this task were only comparable to our other studies for its round 1 data (predictions of round 2 onwards were contaminated by learning about round 1 performance). Third, the other individual they had to predict for was a “randomly selected previous participant” (not a typical person or the person next to them). Accordingly, the actual score of the other was the score of a randomly selected study participant. Thus, this means that this data is not fully comparable to our other studies.

Nevertheless, despite these conceptual differences, the Placement Belief x Belief Correctness interaction replicated across all observations of all 18 rounds<sup>2</sup> (pretask predictions:  $F(1, 1431) = 834.13, p < .001, \eta_p^2 = .37$ , and posttask predictions:  $F(1, 1427) = 104.84, p < .001, \eta_p^2 = .07$ ) and across first round observations only (pretask predictions:  $F(1, 78) = 78.62, p < .001, \eta_p^2 = .50$ , and posttask predictions:  $F(1, 78) = 15.49, p < .001, \eta_p^2 = .17$ ). Importantly, the pattern of means conceptually replicated the results of our prior studies as summarized in the meta-analysis (cf. Table 2), while participants with erroneous WTA beliefs were more clearly

---

<sup>2</sup> Note, one observation (round 18 score of participant #9) was missing.

pessimistic about their absolute performance in this context, in which they compared themselves to a randomly selected prior participant and over multiple rounds. Note, the pattern from round 1 data needs to be interpreted carefully because of low cell sizes given the low number of overall observations ( $N = 82$ ).

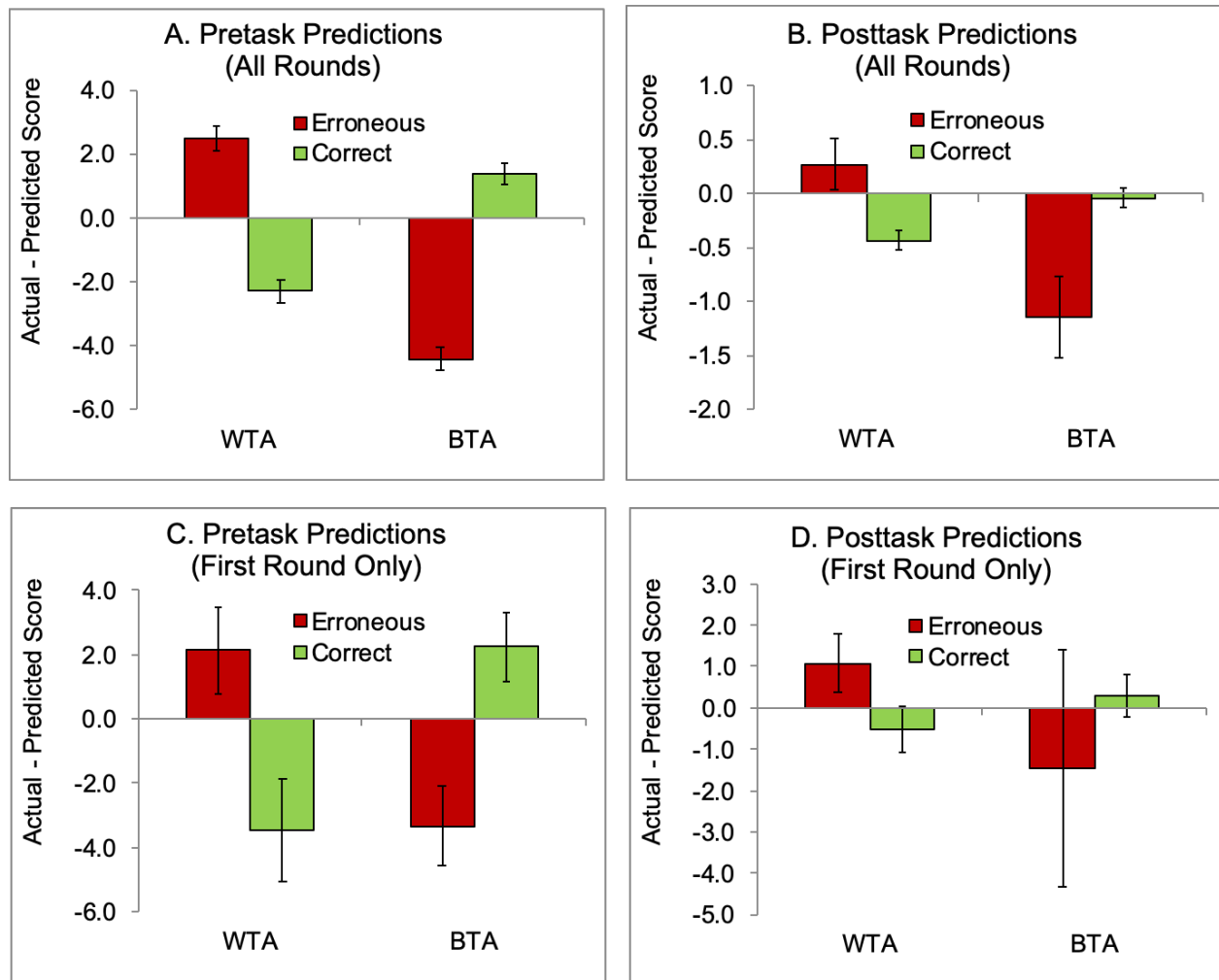


Figure S5. Misestimation of Own Absolute Performance (Reexamination Moore & Healy, 2008).

Positive values mean performing *better* than predicted (*underestimation* of performance), and negative values mean performing *worse* than predicted (*overestimation* of performance). Error bars represent 95% confidence intervals.

## REFERENCES

- Baumgartner, H., Pieters, R., & Bagozzi, R. P. (2008). Future-oriented emotions: conceptualization and behavioral effects. *European Journal of Social Psychology*, 38(4), 685-696. doi: 10.1002/ejsp.467
- Engeler, I., & Raghurir, P. (2018). Decomposing the Cross-Sex Misprediction Bias of Dating Behaviors: Do Men Overestimate or Women Underreport their Sexual Intentions? *Journal of Personality and Social Psychology*, 114(1), 95-105. doi:10.1037/pspi0000105
- Leone, L., Perugini, M., & Bagozzi, R. (2005). Emotions and decision making: Regulatory focus moderates the influence of anticipated emotions on action evaluations. *Cognition & Emotion*, 19(8), 1175-1198. doi: 10.1080/02699930500203203
- Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502-517. doi:10.1037/0033-295x.115.2.502
- Moore, D. A., & Small, D. A. (2007). Error and bias in comparative judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology*, 92(6), 972-989. doi:10.1037/0022-3514.92.6.972
- Oettingen, G., & Mayer, D. (2002). The motivating function of thinking about the future: Expectations versus fantasies. *Journal of Personality and Social Psychology*, 83(5), 1198-1212. doi: 10.1037//0022-3514.83.5.1198
- Oettingen, G., Pak, H. J., & Schnetter, K. (2001). Self-regulation of goal-setting: turning free fantasies about the future into binding goals. *Journal of Personality and Social Psychology*, 80(5), 736-753. doi: 10.1037/10022-3514.80.5.736



- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315-341. doi: 10.1007/s10648-006-9029-9
- Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin*, 85(1), 185-193. doi:10.1037//0033-2909.85.1.185