Heritability of different types of overconfidence

Jacob Dooley (University of Sydney)

Nathan Kettlewell* (University of Technology Sydney)

Agnieszka Tymula (University of Sydney)

Abstract:

Incorrect estimation of own absolute and relative abilities is common and can have detrimental effects on a person's educational, social, employment, and financial outcomes. It is not yet fully understood from where interpersonal differences in overconfidence emerge. In this paper, we estimate the heritability of two types of overconfidence, overestimation and overplacement, in a sample of 1120 twins. We find that the genetic heritability of both types of overconfidence is about 19% and that most of the interindividual variation in overconfidence is due to individual-specific environmental factors.

Keywords: heritability, overconfidence, genetics, twin study

JEL: D91; Z13

Funding: We received funding support from the University of Technology Sydney and from the Australian Research Council's Centre of Excellence for Children and Families over the Life Course (Project ID CE140100027 and CE200100025).

*Corresponding Author: Nathan Kettlewell (Nathan.Kettlewell@uts.edu.au)

1. Introduction

Overconfidence (and underconfidence) is a miscalibrated belief about own absolute or relative performance. Researchers investigated the consequences of and individual differences in overconfidence. For example, men tend to be more overconfident than women leading to male overrepresentation in leadership positions (Reuben et al., 2012). Overconfidence can have perverse effects including poor financial decision-making (Camerer & Lovallo, 1999) and academic complacency (Vancouver et al., 2008).

We focus on the foundations of overconfidence by asking: 'to what extent are interpersonal differences in overconfidence explained by differences in genes and in nurture?'. Only two twin studies estimate the genetic heritability of overconfidence. Cesarini et al. (2009) find that 16%-34% of interpersonal variation in the tendency to overestimate own relative performance (overplacement) is due to differences in genes among adult Swedish twins. Vogt et al. (2021) attribute 18%-28% of interpersonal differences in overestimation (measured as a residual between subjective belief in and objective performance) to genes in a sample of children aged 7-15 years in Texas. The remaining variation in overconfidence in both studies was mainly due to unique personal circumstances.

We extend earlier work by measuring different types of overconfidence in one sample. We consider both overestimation, the overconfidence in own absolute performance, and overplacement, the overconfidence in own relative performance. Overestiamtion and overplacement are conceptually (and empirically) distinct (Moore & Healy, 2008), with overestimation relying on imperfect information about own performance and overplacement additionally relying on imperfect beliefs about other's performance. We estimate the extent to which these distinct traits are explained by genes and the extent to which their correlation is explained by shared genetic influences. The potential interaction between genes and the common social environment means that the heritability of behavioral traits may vary across social contexts. Existing twin studies provide only two data points: for adult Swedes and for children in Texas. Here, we study a different social context (Australia) and check whether previous findings generalize.

2. Data

We analyze the Australian Twins Economic Preferences Survey (ATEPS) (Kettlewell & Tymula 2021). 560 twin pairs (ages 18-66, mean 44.7) completed the online survey in 2020-21. 401 pairs are monozygotic (identical) and 159 are dizygotic (fraternal)[1]. 83% of the sample is female and 59% have a university degree. Further demographic details are in Tables A1-A2.

In one task, participants undertook a cognitive challenge in which they were incentivized to solve ten puzzles from the matrix reasoning item bank developed by Chierchia et al. (2019). Participants had 30 seconds to select the correct image from four available options to complete a sequence (Figure A1 presents an example). They completed a practice round before starting the task. Puzzles had different difficulty levels, with an expected average score of 6/10. Participants earned AU$2 per correct answer if this task was chosen for payment.

Upon completing the task, participants reported "How many of the ten puzzles do you think you got right?" and "Where do you think you will rank in the puzzle task compared to other twins in this study, out of 100?". The differences between perceived and actual score and rank are our measures of overconfidence (overestimation and overplacement).

Participants correctly answered 5.0 (s.d.=1.9) puzzles on average. They are slightly underconfident about their score (μ=-0.57 s.d. = 2.23) but overconfident about their rank (μ=13.42 s.d. = 39.96). The distributions of both types of overconfidence look normal (Figure A2) and overestimation and overplacement are slightly correlated (r=0.20).

3. Methodology

To decompose the variation in overconfidence into genetic and environmental effects, we exploit that monozygotic (MZ) twins are genetically identical, while dizygotic (DZ) twins share 50% of the genes on average (assuming no assortative mating). The variance of a trait Y is $\sigma_Y^2 = \sigma_a^2 + \sigma_d^2 + \sigma_c^2 + \sigma_e^2$ where the independent components are additive genetic effects (σ_a^2), dominance genetic effects (where expression of genes depends on interactions at particular

---

[1] 42 are mixed sex and so we control for sex in our analysis.

loci) (σ_d^2), environmental effects common to siblings (e.g., from shared parenting) (σ_c^2), and environmental effects that are unique (σ_e^2), which include noise. Identification comes from the assumption that the common family environment contributes to overall variance the same for MZ and DZ twins. Because additive and dominance genetic effects cannot be separately identified, $\sigma_d^2 = 0$ is frequently imposed. Under these assumptions, the correlation of Y between MZ twins (r_{MZ}) captures $(\sigma_a^2 + \sigma_c^2)/\sigma_Y^2$ and between DZ twins (r_{DZ}) $(0.5\sigma_a^2 + \sigma_c^2)/\sigma_Y^2$. These correlations can be manipulated to estimate narrow sense heritability: $h^2 = \sigma_a^2/\sigma_Y^2 = 2 * (r_{MZ} - r_{DZ})$.

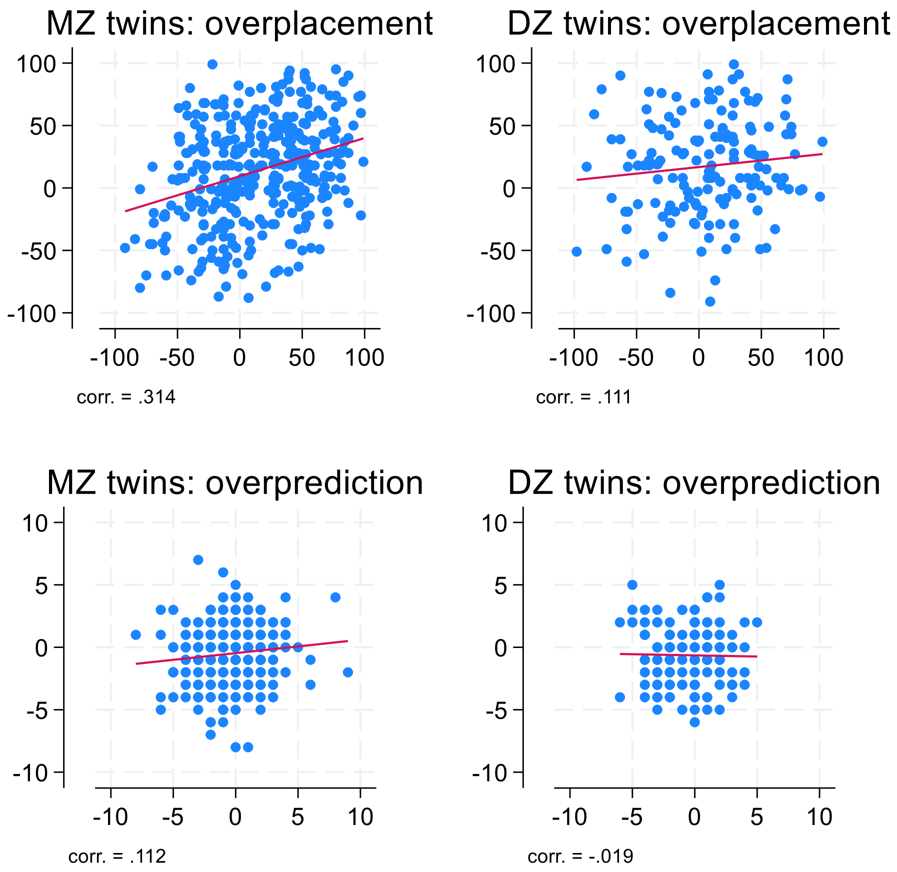
In practice, researchers assume normality for the variance components and use structural equation modeling (SEM) to estimate variance shares (Neale & Cardon, 1992). The SEM approach is appealing because it ensures variance shares sum to one, it can accommodate controls and can be extended to multivariate decompositions, here allowing to estimate the degree to which overlapping genes explain the correlation between different overconfidence types. We follow the SEM tradition and estimate the bivariate AC(D)E-Cholesky model, using the *umx* package for R (Bates et al., 2019). We report our estimates as shares of explained variance.

We estimate four versions of the model and compare their fit. Our baseline model imposes $\sigma_d^2 = 0$ (ACE), most common in the literature. Next, we impose $\sigma_d^2 = \sigma_c^2 = 0$ (AE), then $\sigma_a^2 = \sigma_d^2 = 0$ (CE), and finally $\sigma_c^2 = 0$ (ADE). Like Cesarini et al. (2009), we estimate our models with and without controlling for individual puzzle task score, since very low/high scores constrain the ability to reveal under and overconfidence. We use non-parametric bootstrap clustered at the twin pair level (999 replications) to estimate standard errors and confidence intervals.

4. Results

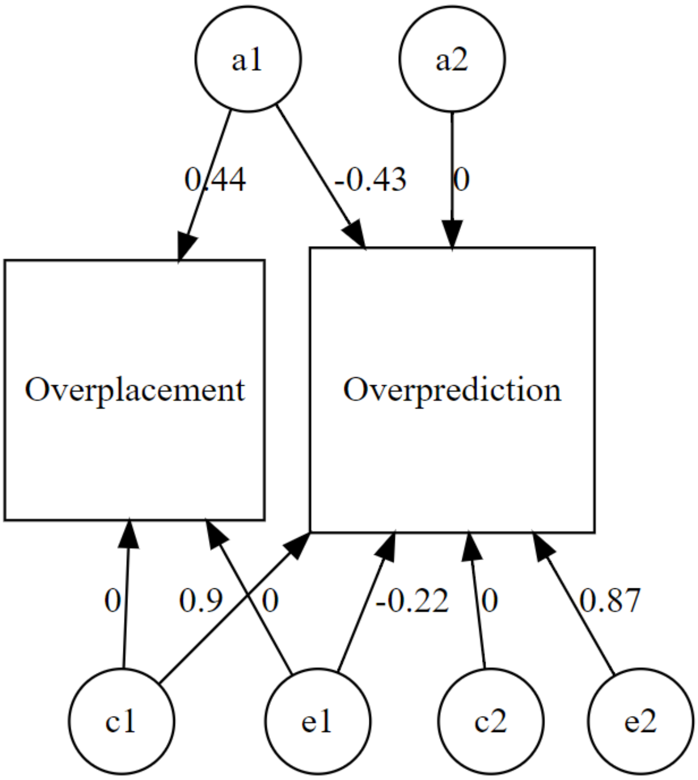
First, we compare pairwise correlations in overconfidence for MZ and DZ twins (Figure 1). For overestimation, $r_{MZ} = 0.11$ (s.e. = 0.049) and $r_{DZ} = -0.02$ (s.e. = 0.089). The stronger correlation for MZ twins hints at the importance of genes. The lack of correlation for DZ twins suggests dominance genetic effects might be important (i.e., genes act multiplicatively); however, given fewer DZ twins this may reflect imprecision. Overprediction correlations are stronger, $r_{MZ} = 0.31$ (s.e. = 0.047) and $r_{DZ} = 0.11$ (s.e. = 0.080), pointing to genes accounting for part of the interpersonal differences.

Figure 1: Overconfidence correlations



Next, we formally estimate the variance shares using SEM (Table 1). The SEM path diagram (Figure 2) shows the ACE model standardized factor loadings, including the covariances. When adjusting for the task score (Panel A), we estimate genetic heritability of 19% [CI=0.11%, 26%] for overplacement and 19% [12%, 26%] for overprediction in the ACE specification (AE is almost identical). AE model fits best and ADE fits slightly better than ACE. The implied heritability is similar across model specifications with little evidence for the importance of common family environment. There is strong negative correlation between genes that explain overplacement and overestimation (Figure 2). Following Kuntsi et al. (2004), we calculate the genetic correlation coefficient = -1.00 [-0.61, -1] implying a near perfect relationship (for unadjusted overconfidence corr. = -0.85 [-0.04, -1]). This implies that genes that increase the tendency for overplacement decrease the tendency for overprediction.

Figure 2: SEM path diagram (standardized coefficients)



Notes: Standardized factor loading for ACE model controlling for score in puzzle task.

Table 1: AC(D)E-Cholesky decomposition estimates

		A.1: Overplacement (adjusted)		
	ACE	AE	CE	ADE
$\hat{a}^2$	0.19 (0.11, 0.26)	0.19 (0.12, 0.26)		0.00 (0.00, 0.24)
$\hat{d}^2$				0.21 (0.10, 0.29)
$\hat{c}^2$	0.01 (0.00, 0.14)		0.14 (0.07, 0.22)	
$\hat{e}^2$	0.81 (0.74, 0.88)	0.81 (0.74, 0.88)	0.81 (0.78, 0.93)	0.79 (0.72, 0.87)
$\hat{a}^2 + d^2$				0.21 (0.13, 0.28)
		A.2: Overprediction (adjusted)		
$\hat{a}^2$	0.19 (0.12, 0.26)	0.19 (0.11, 0.26)		0.00 (0.00, 0.49)
$\hat{d}^2$				0.20 (0.10, 0.28)
$\hat{c}^2$	0.01 (0.00, 0.15)		0.14 (0.09, 0.22)	
$\hat{e}^2$	0.81 (0.74, 0.88)	0.81 (0.74, 0.89)	0.86 (0.78, 0.91)	0.80 (0.73, 0.89)
$\hat{a}^2 + d^2$				0.20 (0.11, 0.27)
AIC	4608.620	4602.624	4611.731	4605.124
χ^2 p-val	-	1.00	0.028	-
		B.1: Overplacement (unadjusted)		
$\hat{a}^2$	0.25 (0.13, 0.33)	0.29 (0.23, 0.36)		0.09 (0.00, 0.28)
$\hat{d}^2$				0.21 (0.05, 0.32)
$\hat{c}^2$	0.04 (0.00, 0.16)		0.24 (0.18, 0.31)	
$\hat{e}^2$	0.71 (0.64, 0.77)	0.71 (0.64, 0.77)	0.76 (0.69, 0.82)	0.70 (0.64, 0.77)
$\hat{a}^2 + d^2$				

		B.2. Overprediction (unadjusted)		
$\hat{a}^2$	0.05 (0.01, 0.12)	0.10 (0.03, 0.18)		0.04 (0.00, 0.56)
$\hat{d}^2$				0.07 (0.02, 0.14)
$\hat{c}^2$	0.04 (0.00, 0.11)		0.07 (0.01, 0.14)	
$\hat{e}^2$	0.71 (0.64, 0.77)	0.90 (0.82, 0.97)	0.93 (0.86, 0.99)	0.89 (0.82, 0.96)
$\hat{a}^2 + d^2$				0.11 (0.04, 0.18)
AIC	6231.450	6225.617	6238.001	6227.943
$\Delta\chi^2$ p-val	-	0.983	0.006	-

Notes: SEM estimates for the additive genetic effects ($\hat{a}^2$), dominance genetic effects ($\hat{d}^2$), common environment ($\hat{c}^2$), and unique environment ($\hat{e}^2$). Adjusted estimates control for the puzzle task score; all estimates control for sex. $\Delta\chi^2$ p-val tests the improvement in log-likelihood relative to ACE. Nonparametric bootstrap 95%-level confidence intervals clustered at the twin pair level in parentheses.

Without controlling for the task score (Panel B), the genetic variance estimates for overplacement are larger (29% [23%,36%] in the best-fitting AE model), which is also the case in Cesarini et al. (2009) (34% unadjusted, 16% adjusted). However, our estimates for overestimation become smaller (10% [3%,18%]). In either case, conclusions are qualitatively similar – unique experiences explain most variation, but genes are a non-trivial factor.

5. Discussion

We provide new estimates for the extent to which interpersonal differences in overconfidence are explained by genes versus nurture in an Australian adult twins sample. Genetic variation can account for 19%-29% of variation in overplacement, in line with Cesarini et al. (2009). We estimate the heritability of overestimation at 10%-19%. The remaining variation for both measures is due to unique experiences.

Overconfidence has been implicated in various achievement gaps, including worse educational and financial outcomes and gender gaps in the labor market. This begs the question, why are some people more (over)confident than others, and can we change this? Our estimates suggest overconfidence is largely due to idiosyncratic circumstances. The relatively modest pressure from genetic forces could mean this trait is quite malleable (although it bears mentioning genes are not deterministic and merely increase the chances of certain behaviors). We find a negative genetic correlation between overplacement and overprediction. This estimate is fairly imprecise and we caution against overinterpretation; nonetheless, the idea that the genes that drive overplacement might also nudge us away from overestimation is intriguing and could have practical implications. For example, researchers are increasingly using polygenetic scores to proxy for different traits. Our results imply that if a score is developed for overconfidence, researchers should be very careful in understanding what type of overconfidence the score reflects before applying it to their work, to avoid misinterpreting results.

References

Bates, T. C., Maes, H., & Neale, M. C. (2019). umx: twin and path-based structural equation modeling in R. *Twin Research and Human Genetics*, 22(1), 27-41.

Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89(1), 306-318.

Cesarini, D., Lichtenstein, P., Johannesson, M., & Wallace, B. (2009). Heritability of overconfidence. *Journal of the European Economic Association*, 7(2-3), 617-627.

Chierchia, G., Fuhrmann, D., Knoll, L. J., Pi-Sunyer, B. P., Sakhardande, A. L., & Blakemore, S. J. (2019). The matrix reasoning item bank (MaRs-IB): Novel, open-access abstract reasoning items for adolescents and adults. *Royal Society Open Science*, 6(10), 190-232.

Kettlewell, N., & Tymula, A. (2021). The Australian Twins Economic Preferences Survey. *Twin Research and Human Genetics*, 24(6), 365-370.

Kuntsi, J., Eley, T. C., Taylor, A., Hughes, C., Asherson, P., Caspi, A., & Moffitt, T. E. (2004). Co-occurrence of ADHD and low IQ has genetic origins. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 124(1), 41-47.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502-517.

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families* (1st ed. 1992.). Springer Science Business Media, B.V. https://doi.org/10.1007/978-94-015-8018-2.

Reuben, E., Rey-Biel, P., Sapienza, P., & Zingales, L. (2012). The emergence of male leadership in competitive environments. *Journal of Economic Behavior & Organization*, 83(1), 111-117.

Vancouver, J. B., More, K. M., & Yoder, R. J. (2008). Self-efficacy and resource allocation: support for a nonmonotonic, discontinuous model. *Journal of Applied Psychology*, 93(1), 35-47.

Vogt, R. L., Zheng, A., Briley, D. A., Malanchini, M., Harden, K. P., & Tucker-Drob, E. M. (2022). Genetic and Environmental Factors of Non-Ability-Based Confidence. *Social Psychological and Personality Science*, 13(3), 734-746.

Online Appendix for “Heritability of different types of overconfidence”

Table A1: Descriptive statistics

Variable	Mean MZ	Mean DZ	Difference	P-value
Age	44.034	46.289	2.256	0.058
Male	0.142	0.245	0.103	0.002
Australia born	0.869	0.903	0.033	0.245
Lives in a city	0.648	0.659	0.012	0.759
Married/defacto	0.653	0.688	0.035	0.304
Household size	4.513	4.438	-0.075	0.577
Num. dep. children	1.89	2.016	0.126	0.289
University degree	0.589	0.597	0.009	0.816
Employed	0.859	0.836	-0.023	0.409
Retired	0.08	0.085	0.005	0.822
Income (weekly)	1256.9	1321.909	65.009	0.246
Financial security	3.158	3.176	0.018	0.748
Long-term health condition	0.218	0.189	-0.029	0.316

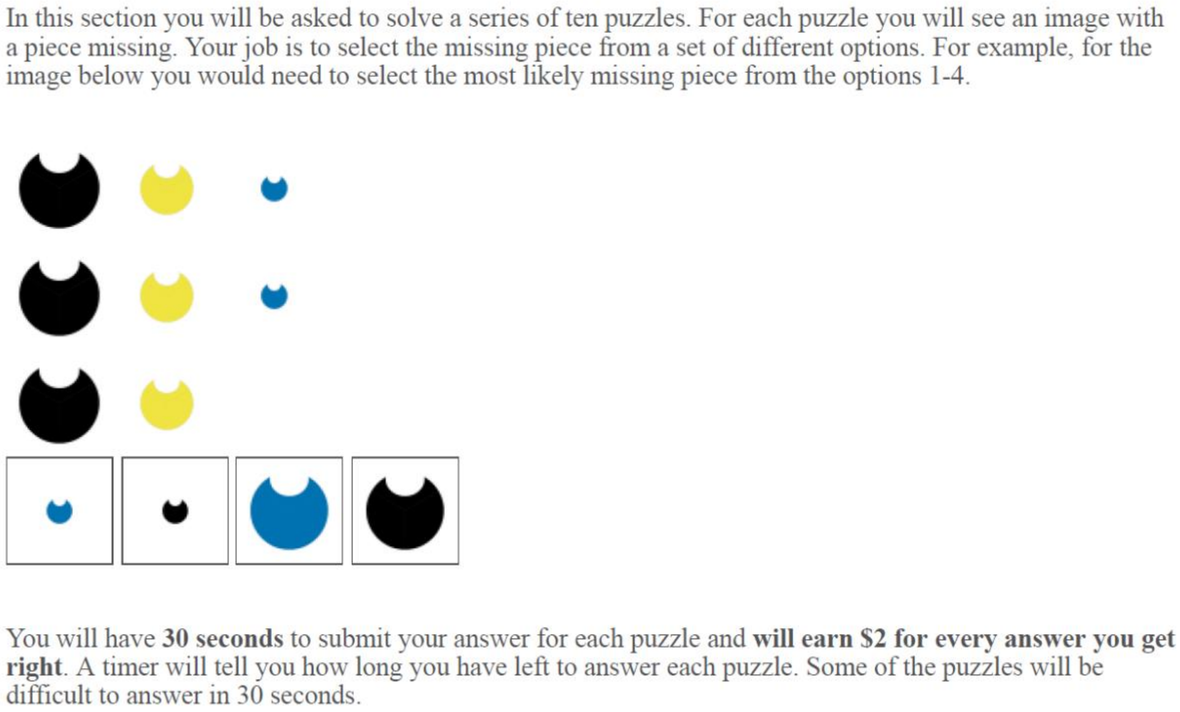
Notes: Calculated from non-missing values from a full sample of 802 monozygotic twins and 318 dizygotic twins. Clustered (twin pair level) standard errors are used to calculate p-values.

Table A2: Variable definitions

Variable	Definition	MZ obs.	DZ obs.
Age	Age at last birthday	802	318
Male	= 1 if male	802	318
Australia born	= 1 if born in Australia	802	318
Lives in a city	= 1 if currently live in a major city (Sydney, Melbourne, Brisbane, Adelaide, Perth, Canberra)	795	317
Married/defacto	= 1 if married or in a defacto relationship	800	314
Household size	How many people live in your household	799	315
Num. dep. children	Number of dependent children	766	310
University degree	= 1 if highest level of education obtained is a university degree	802	318
Employed	= 1 if worked any time in the last 7 days or if had a job but did not work in the last 7 days due to holidays, sickness or any other reason	802	318
Retired	= 1 if currently retired from the workforce	802	318
Income (weekly)	Average usual weekly own income in the last month using midpoint value for the following categories: \$1-\$149, \$150-\$299, \$300-\$399, \$400-\$499, \$500-\$649, \$650-\$799, \$800-\$999, \$1,000-\$1,249, \$1,250-\$1,499, \$1,500-\$1,749, \$1,750-\$1,999, \$2,000-\$2,999, \$3,000 or more (coded as \$3000). Negative or nil coded as missing.	692	275
Financial security	Given your current needs and financial responsibility, would you say that you and your family are: = 1 if Poor, = 2 if Just getting	802	318

	along, = 3 if Comfortable, = 4 if Very comfortable, = 5 if Prosperous.		
Long-term health condition	= 1 if has a long-term health condition, impairment or disability that has lasted more than 6 months	800	318

Figure A1: Instructions and example of cognitive task



In this section you will be asked to solve a series of ten puzzles. For each puzzle you will see an image with a piece missing. Your job is to select the missing piece from a set of different options. For example, for the image below you would need to select the most likely missing piece from the options 1-4.

You will have **30 seconds** to submit your answer for each puzzle and **will earn $2 for every answer you get right**. A timer will tell you how long you have left to answer each puzzle. Some of the puzzles will be difficult to answer in 30 seconds.

Notes: The full survey, with each puzzle, can be downloaded at https://dataverse.ada.edu.au/dataset.xhtml?persistentId=doi:10.26193/TTQEBQ

Figure A2: Distributions of overconfidence measures