Supplemental Materials

Analyses for Correct Reasoners

Confidence

Consistent with De Neys et al. (2013), correct responders were less confident in their correct standard responses ($M = .946$, $SD = .138$) than in their control responses ($M = .981$, $SD = .066$), $F(1, 152) = 10.252$, $p = .002$, $\eta_p^2 = .06$.

Opinion

For correct respondents' opinions of others' ability to answer correctly, there was an effect of subject group (MTurk 1st sample: $M = .710$, $SD = .261$; MTurk 2nd sample: $M = .782$, $SD = .196$; UCB: $M = .788$, $SD = .211$), $F(2, 152) = 3.638$, $p = .029$, $\eta_p^2 = .05$, reflecting lower opinions by the first MTurk sample than by the other groups. Importantly, there was a main effect of condition such that correct respondents reported significantly less confidence in other reasoners' standard responses ($M = .667$, $SD = .220$) than in their control responses ($M = .845$, $SD = .197$), $F(1, 152) = 66.185$, $p < .001$, $\eta_p^2 = .30$.

Memory

For the recall response, a mixed effects logistic regression (with subject as the random variable) was conducted due to the dichotomous dependent variable (with or without "more than"). The effect of condition was significant such that correct reasoners usually recalled the standard problem, but not the control, as containing "more than" (standard: $M = .936$, $SD = .246$, control: $M = .073$, $SD = .262$), with this effect of condition significant, $b = 5.01$, odds ratio (*OR*) $= 149.31$, $\chi^2 = 91.28$, $p < .001$, 95% confidence interval (*CI*) [5.74, 72.75].

In terms of the recognition response, a mixed effects logistic regression (with subject as the random variable) was also conducted due to the dichotomous dependent variable (with or

without "more than"). Correct reasoners recognized the standard problem, but not the control, as containing "more than" (standard: $M = .961$, $SD = .194$, control: $M = .103$, $SD = .305$), with this effect of condition significant, $b = 5.15$, odds ratio (OR) $= 171.623$, $\chi^2 = 122.74$, $p < .001$, 95% confidence interval (CI) [69.07, 426.49].

Response Time

There was a main effect of subject group on response time for correct respondents (MTurk 1st sample: $M = 31.143$, $SD = 48.942$; MTurk 2nd sample: $M = 39.594$, $SD = 60.110$; UCB: $M = 49.268$, $SD = 60.164$), $F(2, 152) = 4.536$, $p = .012$, $\eta_p^2 = .06$, reflecting slower responses by UCB than by MTurk groups. Importantly, there was a main effect of condition such that correct participants responded faster to the control question ($M = 22.520$, $SD = 25.990$) than to the standard question ($M = 57.191$, $SD = 72.538$), $F(1, 152) = 78.611$, $p < .001$, $\eta_p^2 = .34$. This finding can be explained simply by the additional arithmetic operation required by the standard question. Condition and subject group significantly interacted (MTurk 1st sample: standard: $M = 39.111$, $SD = 62.343$; control: $M = 23.175$, $SD = 28.710$; MTurk 2nd sample: standard: $M = 55.680$, $SD = 79.460$; control: $M = 23.508$, $SD = 21.177$; UCB: standard: $M = 77.938$, $SD = 69.340$; control: $M = 20.597$, $SD = 28.814$), $F(2, 152) = 15.173$, $p < .001$, $\eta_p^2 = .17$, such that the difference in response time between standard and control was greater for the UCB sample than it was for the MTurk samples.

Most important was a negative correlation between the difference in response time and the difference in response confidence for correct reasoners, $r(153) = -.181$, $p = .024$, such that longer times for the standard relative to the control condition were associated with lower confidence for the standard relative to the control condition.

Significant Effects Involving Subject Group for Incorrect Reasoners

Opinion

There was a main effect of subject group (MTurk 1st sample: $M = .775$, $SD = .214$;

MTurk 2nd sample: $M = .828$, $SD = .230$; UCB: $M = .859$, $SD = .153$), $F(2, 224) = 4.768$, $p =$

.009, $\eta_p^2 = .04$, reflecting lower opinions by the first MTurk sample than by the UCB sample.

Also, condition and subject group significantly interacted (MTurk 1st sample: Standard: $M =$

.785, $SD = .192$; Control: $M = .765$, $SD = .235$; MTurk 2nd sample: Standard: $M = .795$, $SD =$

.241; Control: $M = .862$, $SD = .216$; UCB: Standard: $M = .831$, $SD = .169$; Control: $M = .886$, SD

$= .132$), $F(2, 224) = 4.067$, $p = .018$, $\eta_p^2 = .04$, such that the lower opinion judgments for the

standard than for the control was found for the UCB and the second MTurk sample but not for

the first MTurk sample.

Simple Math Problems (from Hathorn & Healy, 2015, 2016, and Hoover & Healy, 2017)

Added to the Task for MTurk Sample 2 and the UCB Sample

(with observed mean proportion correct and standard deviation)

1.   Arthur baked 115 muffins, which was 17 more muffins than Ann. How many muffins did

        Ann bake? (Answer: 98 muffins; Proportion correct: $M = .941$, $SD = .235$)

2.   In the summertime, you can earn $4 a day by cutting the grass. How many days will it

        take you to earn $84? (Answer: 21 days; Proportion correct: $M = .969$, $SD = .174$)

3.   Willy has 92 crayons. Lucy has 73 crayons. How many more crayons does Willy have

        than Lucy? (Answer: 19 crayons; Proportion correct: $M = .977$, $SD = .152$)

4.   The turtle at the zoo weighs 45 pounds. It is five times heavier than the baby turtle. How

        much does the baby turtle weigh? (Answer: 9 pounds; Proportion correct: $M = .910$, SD

        $= .287$)

5.   Suppose you want to buy three loaves of bread that cost $1.50 each and a jar of peanut

        butter that costs $4. A jar of jelly is $2.75, but you don't need any jelly. You have $10.

        How much money will you have left over? (Answer: $1.50; Proportion correct: $M =$

        .785, $SD = .412$)

6.   The star running back on our football team got most of his total yardage running. The rest

        was catching passes. He caught passes for 60 yards. His total yardage was 150 yards. The

        running back for the other team got 200 yards. How many yards did the star running back

        on our football team get running? (Answer: 90 yards; Proportion correct: $M = .734$, $SD =$

        .443)

7.   The average temperature in Lincoln in July is 85 degrees. Last Wednesday, it was 90

        degrees. Today it was 15 degrees cooler than last Wednesday. What was the temperature

        today? (Answer: 75 degrees; Proportion correct: $M = .918$, $SD = .275$)

8.   Joanna has 13 more books in her backpack than Sophie. If Sophie has 11 books, how

        many does Joanna have? (Answer: 24 books; Proportion correct: $M = .930$, $SD = .256$)

9.   Consider the following equations. Use the substitution method to solve for the value of X

        X + 3Y = 13
            Y = 6 - 2X
        Solve for X.
        X =
(Answer: X = 1; Proportion correct: $M = .578$, $SD = .495$)