Belief-based and analytic processing in transitive inference: Further evidence for the importance of premise integration

Glenda Andrews

&

Mandy Mihelic

School of Applied Psychology

&

Behavioural Basis of Health Research Program

Griffith Health Institute

Griffith University

Gold Coast Campus, 4222

Australia

Running head: PREMISE INTEGRATION IN TRANSITIVE INFERENCE

Corresponding author: Glenda Andrews, Email: g.andrews@griffith.edu.au

Phone: + 617 5552 8613. Fax: +617 5552 8291

Mandy Mihelic, Email: mandy.mihelic@griffithuni.edu.au

Abstract

Transitive inference problems were presented in a modified conclusion evaluation procedure. Conclusions were believable or unbelievable and valid, invalid, or indeterminate. The 67 undergraduate students read the premises, evaluated the conclusions (accept, reject, cannot tell), then provided confidence ratings. Fluid intelligence was also assessed. Acceptance of valid conclusions, rejection of invalid conclusions, and cannot tell responses to indeterminate conclusions of non-integrable problems indicated use of analytic processing. Believability effects indicated use of heuristic processing. Fluid intelligence and premise integration ability (non-integrable problems) predicted greater use of analytic processing on valid and invalid problems. Premise integration ability was associated with reduced belief bias on invalid problems. Premise integration ability appears to influence the extent of heuristic versus analytic processing. Confidence was sensitive to the presence of belief-logic conflict. Conflict detection scores reflecting this sensitivity were not associated with analytic processing suggesting that conflict detection occurs automatically and reflects an intuitive logic.

Belief bias in reasoning is the tendency to evaluate conclusions based on beliefs rather than logical form (Evans, 2007). Given the premises, *The car is faster than the bicycle* and *The bicycle is faster than the plane*, accepting the conclusion, *The plane is faster than the car,* indicates belief bias. The logical response is to reject the conclusion.

Most explanations of belief bias invoke dual-process models, in which human reasoning involves two types of processes (Evans, 2007). Analytic processes are deliberate, slow and sequential. They permit abstract and hypothetical thinking, but are cognitively effortful. Heuristic processes occur rapidly and require less effort. They can include retrieval of beliefs from memory. Belief bias is observed when analytic and heuristic processes conflict, and heuristic processing dominates.

Heuristic and analytic processing can be detected using the conclusion evaluation procedure. The premises and conclusion are presented and participants accept or reject the conclusion. Use of analytic and heuristic processes is inferred from acceptance rates for conclusions that differ in validity and believability. Thus there are valid believable (VB), valid unbelievable (VU), invalid believable (IB), and invalid unbelievable (IU) problems. VB and IU problems are called non-conflict problems because heuristic and analytic processes yield the same conclusion. VU and IB problems are called conflict problems, because heuristic and analytic processes yield different conclusions. Acceptance rates are higher for valid than invalid, and for believable than unbelievable conclusions. These effects reflect analytic and heuristic processes, respectively. Believability effects are typically stronger for invalid than valid conclusions (Evans, 2007).

Much of the empirical research has been conducted using categorical syllogisms (e.g., Evans & Curtis-Holmes, 2005; Stupple & Ball, 2008). Belief bias has been less studied in other forms of deductive reasoning such as relational reasoning (Banks, 2013, Roberts & Sykes, 2003) and transitive inference, as examined here (Andrews, 2010).

Transitive inference has the form, if a > b and b > c, then the conclusion a > c follows, where > is an asymmetric transitive relation and a, b and c are the elements. For example, *Tom is taller than Jessica* and *Jessica is taller than Lucy* therefore *Tom is taller than Lucy*. Most people make transitive inferences by integrating the elements into an ordered left-right or top-down array (Prado, VanDer Henst, & Noveck, 2008). Once the premises are integrated, it is simple to confirm by inspecting the array, *Tom - Jessica - Lucy* that *Tom is taller than Lucy*.

Premise integration is central to transitive inference. In a dual-task experiment with probe reaction time as the secondary task, the highest processing load occurred when premises were being integrated (Maybery, Bain, & Halford, 1986). Premise integration is difficult for young children (Andrews & Halford, 1998; Bunch & Andrews, 2012). Integration capacity declines in later adulthood (Andrews & Todd, 2008) and is seriously impaired following lesions to the prefrontal cortex (Waltz, Knowlton, Holyoak, Boone, Mishkin, De Menezes Santos et al., 1999).

Given the importance of premise integration, it is crucial to ensure that it cannot be circumvented. Thayer and Collyer (1978) described a non-integrative, labelling strategy. In the previous example, Tom is mentioned once as the taller person and Lucy as the shorter person. Thus, Tom would be labelled as tall and Lucy as short. The conclusion can be evaluated by referring to these labels, without integrating the premises. Labelling is avoided in five-term problems with four premises, a > b, b > c, c > d, d > e. The conclusion relates elements b and d, each of which occurs as the larger and the smaller element in the premises. This removes the advantage of labelling. Participants must integrate the premises to reach the conclusion.

Andrews (2010) examined heuristic and analytic processing in five-term transitive inference problems. Acceptance rates were higher for believable than unbelievable, and valid than invalid conclusions, consistent with the use of heuristic and analytic processing. Premise integration difficulty was manipulated by varying the time allowed for premise encoding and integration (Experiment 1); by presenting premises in a scrambled top-down order versus the order that was consistent with the elements in the five-term series (i.e., a > b, b > c, c > d, d > e), which allowed a simpler concatenation strategy (Experiment 2); and by varying lexical markedness (Experiment 3). Premise integration was expected to be easier when unmarked terms (e.g., *longer*) were used in four premises than when unmarked terms and marked terms (e.g., *shorter)* were used in two premises each. In all experiments, more heuristic processing and less analytic processing occurred in conditions where premise integration was more difficult.

The current study extended this research. One purpose was to determine whether the believability effects would be replicated when three response options (accept, reject, cannot tell) were available. To accommodate this, non-integrable problems were presented as well as VB, VU, IB and IU problems. Non-integrable problems had four premises, a > b, b > c, d > e, e > f, and conclusions of the form, b > d or d > b. Notice that elements a, b and c can be integrated to form, a > b > c. Elements d, e, and f can be integrated to form, d > e > f. However, a > b > c cannot be integrated with d > e > f, so the relation between b and d is indeterminate. Non-integrable problems assess the capacity to recognize when integration is not possible. The (analytically) correct response is cannot tell.

If participants who give non-normative responses for VU and IB problems do not rely on heuristic processing, they would be expected to respond “cannot tell” and this would weaken the believability effect on acceptance rates. With three response options, belief-based responses on conflict problems can be more clearly identified. Rejecting VU conclusions or accepting IB conclusions are belief-based errors. Cannot tell responses are also incorrect, but they cannot be described as belief-based errors. To the extent that the errors on conflict problems are due to heuristic processing, more belief-based errors than cannot tell errors would be expected.

Specific dual-process models have been proposed to account for belief bias in categorical syllogisms. Stupple and Ball (2008) characterized these as belief-first, reasoning-first or parallel-process models, according to the sequence in which heuristic and analytic processing occur. In belief-first models, believable conclusions are initially accepted and unbelievable conclusions are initially rejected. Belief-based responses are either supported or inhibited by subsequent analytic processing, although not all reasoners engage the analytic route (Stupple Ball, Evans, & Kamal-Smith, 2011). In reasoning-first models, people first attempt to reason analytically, but they fall back on heuristic processing if analytic processing fails (Stupple & Ball, 2008). In parallel-process models, heuristic processing and analytic processing proceed simultaneously and each produces a response. When the two responses differ, the conflict must be resolved (Sloman, 1996).

A key issue for dual process models of belief bias is how conflict is detected and resolved. This requires that heuristic and analytic processing overlap in time or at least that their outputs are simultaneously available. This is somewhat problematic for serial models because both processes are not always engaged. However, parallel models have been criticized as wasteful of scarce cognitive resources. Both processes are always engaged, even though heuristic processing would often suffice. The logical intuition model (De Neys, 2012) resolves this by proposing an intuitive logical process that is distinct from both deliberate (analytic) processing and intuitive heuristic processing. Intuitive logical and heuristic processes are automatically engaged from the outset. Analytic processing is triggered by conflict between the two intuitive processes. Conflict detection occurs automatically, but analytic processing is needed to resolve the conflict.

Confidence ratings provide a potential indicator of conflict-detection. Andrews

(2010) found that participants' confidence in their evaluations was lower for conflict than non-conflict problems indicating that confidence was sensitive to the belief-logic conflict. However it is unclear whether conflict detection occurs automatically as proposed in the logical intuition model (De Neys, 2012), or whether it is a more deliberate process. In the current study, conflict detection scores were computed and potential links with heuristic and analytic processing were examined. If conflict detection is deliberate, then significant associations with analytic processing would be expected.

Individual differences in heuristic and analytic processing were also examined. Fluid intelligence is the ability to adapt to new situations and understand relationships. Dual-process approaches claim that intelligence is related to analytic processing, but is relatively independent of heuristic processing (Stanovitch & West, 2008). Empirical findings are supportive. Intelligence was positively correlated with analytic (not heuristic) processing in categorical syllogisms (Newstead, Handley, Harley, Wright & Farrelly, 2004). In Andrews' (2010) research on transitive inference, higher fluid intelligence and better premise integration ability (assessed separately) predicted analytic processing. Premise integration capacity involves being able to mentally combine premises and to recognize when integration is not possible. Andrews' (2010) premise integration task required the former, whereas the non-integrable problems in the current study require the latter. Premise integration capacity and fluid intelligence should predict analytic processing in valid and invalid problems.

Method

Participants

The participants were 67 students (15 males, 52 females) enrolled in psychology courses at Griffith University, Australia. Their mean age was 25.86 years (*SD* = 11.28). All were native speakers of English. Thirty-three and thirty-four participants received problem sets 1 and 2 respectively.

Materials and Procedure

The transitive inference problems, plausibility ratings task and the fluid intelligence test were administered to small groups of participants.

Transitive inference. In each of two sets, there were 24 problems, four instances of the six problem types. The VB, VU, IB and IU problems were drawn from Andrews (2010). The NB and NU were generated anew. See Table 1 for examples.

Each VB, VU, IB and IU problem had four premises, which when integrated formed an ordered five-term sequence (i.e., a > b > c > d > e). The elements in each problem came from the same semantic category. The premises were expressed using unmarked (e.g., larger) forms in two premises and marked (e.g., smaller) forms in two premises. Valid conclusions were of the form b > d. Invalid conclusions were of the form d > b.

Non-integrable (NB, NU) problems had four premises, a > b; b > c; d > e; e > f and indeterminate conclusions relating b and d terms. Conclusions were believable in NB problems and unbelievable in NU problems.

The problems were presented using PowerPoint software, with images projected onto a large white screen. Instructions were shown on the screen and read aloud by the researcher. Participants were instructed to read the premises, think about how they related to each other, and to assume that all premises were true. When the conclusion appeared, they were to evaluate it in relation to the premises.

Three practice problems with neutral content were presented, one valid, one invalid, and one non-integrable. On each problem, the *ready* signal was presented for 3 seconds, then the premises were displayed in a scrambled top-down order for 30 seconds, followed by the conclusion for 10 seconds, then the *respond now* signal for 10 seconds. Participants recorded their responses (accept, reject, cannot tell) on the sheet provided and rated their confidence in each decision as 0 (low), 1 (medium), or 2 (high). Feedback after each problem encouraged participants to integrate the premises and to recognize when integration was not possible.

The procedure for test problems was similar but the durations were briefer: *ready* (3 seconds); premises (20 seconds); conclusions (5 seconds); *respond now* (5 seconds), and no feedback was provided. Test problems were arranged into four 6-item blocks, each containing one problem of each type. A 30-second break was inserted after each block.

Premises were presented first followed by the conclusion. This procedure encourages forward reasoning from premises, rather than backward reasoning from conclusions as occurs with simultaneous presentation. In research using the production paradigm which requires forward reasoning, belief bias in categorical syllogisms was eliminated (Morley, Evans & Handley, 2004). However, Andrews (2010) observed significant believability effects in transitive inference using sequential presentation. Use of sequential presentation allows comparison with Andrews' findings.

Participants also rated the plausibility of the conclusions of the test problems (presented alone) on a 7-point scale (1 = extremely implausible to 7 = extremely plausible).

Fluid intelligence. Fluid intelligence was assessed using the Culture Fair Test, Scale 2, Form A (Cattell & Cattell, 1973). The four subtests, Series, Classification, Matrices and Conditions, were administered using standard procedures. Subtest scores were summed (max. = 46).

Results and Discussion

Conclusion plausibility.

Single-sample *t*-tests confirmed that rated plausibility of believable conclusions was higher than the neutral midpoint (4) in set 1 (M = 5.39, SD = .65), t (32) = 12.36, $p < .001$, and set 2 (M = 6.01, SD = .64), t (33) = 18.35, $p < .001$. Plausibility of unbelievable conclusions was lower than the neutral midpoint in set 1 (M = 2.20, SD = .75), t (32) = -13.66,

$p < .001$, and set 2 (*M* = 2.99, *SD* = .69), *t* (33) = -8.50, $p < .001$.

Conclusion evaluations.

In the standard conclusion evaluation procedure (Evans, 2007) there are two response options (accept, reject). Acceptance rates are analyzed but rejection rates are not. The current procedure involved three response options. For descriptive purposes we analyzed all three, while recognizing that the measures are not independent.

Preliminary analysis revealed no significant differences between problem sets and no gender differences in fluid intelligence, acceptance, rejection and cannot tell rates, nor the indices derived from them. Figure 1 shows the acceptance, rejection and cannot tell rates.

Acceptance. Acceptance rates were subjected to a 3 × 2 repeated-measures analysis of variance (ANOVA) with validity (valid, invalid, indeterminate) and believability (believable, unbelievable) as independent variables. The validity effect was significant, *F* (2, 132) = 118.42, $p < .001$, partial $\eta^2 = .64$. Acceptance rates were higher for valid (*M* = .71; *SE* = .022) than invalid conclusions (*M* = .26; *SE* = .024), *F* (1, 66) = 142.28, $p < .001$, partial $\eta^2 = .68$. Acceptance rates for invalid and indeterminate (*M* = .26; *SE* = .023) conclusions (for which accept was incorrect) did not differ significantly, partial $\eta^2 = .001$. The believability effect was significant, *F* (1, 66) = 46.93, $p < .001$, partial $\eta^2 = .42$. Acceptance rates were higher for believable (*M* = .50; *SE* = .019) than unbelievable (*M* = .33; *SE* = .017) conclusions.

The overall Validity × Believability interaction was not significant (partial $\eta^2 = .03$). However, the interaction of believability with the contrast between valid versus invalid problems approached significance, *F* (1, 66) = 3.86, $p = .054$, partial $\eta^2 = .06$. The believability effect tended to be larger for invalid (partial $\eta^2 = .38$) than valid (partial $\eta^2 = .12$) problems. The believability effect was significant for non-integrable problems (partial $\eta^2 = .17$). Believable conclusions were (incorrectly) accepted more frequently than unbelievable conclusions.

The believability effects observed on acceptance rates for VB, VU, IB and IU conclusions by Andrews (2010) were replicated despite the possibility that the cannot tell option would eliminate the effect.

Rejection. A 3 (Validity) x 2 (Believability) repeated-measures ANOVA yielded a significant believability effect, F (1, 66) = 37.03, p < .001, partial η^2 = .36. Rejection rates were higher for unbelievable (M = .41; SE = .017) than believable (M = .27; SE = .016) conclusions. The validity effect was significant, F (2, 132) = 93.18, p < .001, partial η^2= .59. Rejection rates were higher for invalid (M = .61; SE = .028) than valid (M = .19; SE = .018) conclusions, F (1, 66) = 127.53, p < .001, partial η^2= .659. Rejection rates for valid and indeterminate conclusions (M = .22; SE = .023) (for which reject was incorrect) did not differ significantly (partial η^2= .023).

The Validity × Believability interaction was significant, F (2, 132) = 5.18, p = .007, partial η^2 = .07. The contrast between valid and invalid problems interacted marginally with believability, F (1, 66) = 3.81, p = .055, partial η^2 = .06. The believability effect tended to be larger on invalid (partial η^2 = .31) than valid (partial η^2 = .15) conclusions. The contrast between indeterminate and valid conclusions did not interact with believability (partial η^2 = .023).

Cannot tell. A 3 (Validity) × 2 (Believability) ANOVA yielded a non-significant believability effect (partial η^2 = .04). The validity effect was significant, F (2, 132) = 96.39, p < .001, partial η^2= .59. Cannot tell rates were higher for indeterminate (M = 52; SE = .037) than for determinate conclusions, F (1, 66) = 21.60, p < .001, partial η^2 = .624. Cannot tell rates for valid (M = .10; SE = .016) and invalid (M = .12; SE = .017) conclusions (for which cannot tell was incorrect) did not differ significantly (partial η^2= .018).

The Validity × Believability interaction was significant, *F* (2, 132) = 3.20, *p* = .004, partial η^2 = .05. The believability effect was significant for indeterminate conclusions, *F* (1, 66) = 5.43, *p* = .023, partial η^2 = .08. Cannot tell rates were higher for conclusions of NU (*M* = .57; *SE* = .043) than NB (*M* = .46; *SE* = .042) problems. Participants found it easier to respond cannot tell when evaluating unbelievable than believable conclusions. The believability effect was not significant for valid (partial η^2 = .00) nor invalid (partial η^2 = .00) conclusions.

In summary, acceptance rates were highest for valid problems, rejection rates were highest for invalid problems, and cannot tell rates were highest for non-integrable problems, providing evidence of analytic processing. Heuristic processing on problems with determinate (valid or invalid) conclusions was evident on acceptance and rejection rates, but not cannot tell rates. For non-integrable problems, it was evident on acceptance and cannot tell rates, but not rejection rates.

There were more belief-based than cannot tell errors on VU problems, *t* (66) = 3.59, *p* = .001, and IB problems, *t* (66) = 5.40, *p* < .001, suggesting that participants recognized when a determinate conclusion was required even though they did not correctly resolve the belief-logic conflict. On non-integrable problems, erroneous acceptance of believable conclusions exceeded erroneous rejection of unbelievable conclusions, *t* (66) = 2.45, *p* = .017.

Confidence.

Confidence ratings are shown in Figure 1. A 2 (Gender) × 3 (Validity) × 2 (Believability) mixed repeated-measures ANOVA yielded a significant gender effect, *F* (1, 65) = 4.04, *p* = .049, partial η^2 = .06. Males (*M* = 1.24; *SE* = .11) reported higher confidence than females (*M* = 0.99; *SE* = .06). Gender did not interact with other variables. The validity effect was significant, *F* (2, 130) = 6.08, *p* = .003, partial η^2 = .09. Confidence was higher for valid (*M* = 1.21; *SE* = .07) than invalid (*M* = 1.09; *SE* = .06) conclusions, *F* (1, 65) = 5.66, *p*

= .02, partial η^2 = .08. Confidence for indeterminate (*M* = 1.04; *SE* = .07) and invalid conclusions did not differ significantly (partial η^2 = .02). The believability effect was not significant (partial η^2 = .00).

Believability interacted with validity, *F* (2, 132) = 21.69, *p* < .001, partial η^2 = .25. Believability interacted with the contrast between valid and invalid problems, *F* (1, 65) = 36.15, *p* < .001, partial η^2 = .36, reflecting higher confidence for no-conflict (VB, IU) (*M* = 1.26; *SE* = .055) than for conflict (VU, IB) problems (*M* = .093, *SE* = .056), *t* (66) = 7.00, *p* < .001, *d* = 0.85. As in Andrews (2010), confidence was sensitive to the belief-logic conflict.

Believability also interacted with the contrast between problems with invalid and indeterminate conclusions, *F* (1, 65) = 24.60, *p* < .001, partial η^2 = .28. Confidence was lower for NU than NB problems, *F* (1, 66) = 4.38, *p* = .04, partial η^2 = .06, but the reverse was the case for invalid problems.

Correlates of heuristic and analytic processing.

Three indices were computed from acceptance rates for valid and invalid problems (Evans & Curtis-Holmes, 2005). The logic index (VB + VU – IB – IU) reflects analytic processing; the belief index (VB + IB – VU – IU) reflects heuristic processing (belief bias); and the interaction index (VU + IB – VB – IU) reflects the extent to which the belief bias is greater on invalid than valid problems. Conflict detection scores reflecting how much higher confidence was for non-conflict than conflict problems were also computed. Cannot tell responses for NB and NU problems were positively correlated, *r* = .45, *p* < .001. They were averaged and used as a measure of premise integration ability.

Premise integration and fluid intelligence were positively correlated with the logic index (Table 2). When entered as predictors in a multiple regression analysis they accounted for 28% of variance in the logic index. Premise integration (8.8%, *p* = .007) and fluid intelligence (6.7%, *p* = .018) accounted for unique variance (Table 3). Conflict detection scores were not significantly correlated with the logic index, fluid intelligence, or premise integration. Conflict detection appears not to involve analytic processing.

The positive correlation between conflict detection and the belief index approached significance (p = .054). Conflict detection did not protect against belief-bias.

Premise integration and conflict detection were negatively correlated with the interaction index. When entered as predictors in a multiple regression analysis, they accounted for 15.7% of variance in the interaction index. Premise integration (5.76%, p = .041) and conflict detection (6.55%, p = .029) accounted for unique variance (Table 3). Better premise integration and conflict detection were associated with smaller interaction indices. Recall that this index reflects the extent to which belief bias is greater for invalid than valid problems. Follow-up analyses showed that premise integration was negatively correlated with belief bias on invalid, r = -.323, p = .008, but not valid problems (p = .444).

Conflict detection was positively correlated with belief bias on valid, r = .355, p = .003, but not invalid problems (p = .81). The correlation with belief bias on valid problems seems to stem from cannot tell responses. Higher conflict detection scores (for valid problems) were associated with lower cannot tell rates on VB problems, r = -.28, p = .02, and tended to be associated with higher cannot tell rates on VU problems, r = .21, p = .09. To the extent that lower cannot tell rates translate to higher acceptance rates, and higher cannot tell rates translate to lower acceptance rates, the belief bias will be larger.

These differential associations with belief bias suggest that whereas premise integration involves analytic processing that is required for conflict resolution, conflict detection does not.

General Discussion

The research extended our understanding of belief bias in three ways. First, it demonstrated that belief bias in transitive inference is robust. It occurred with sequential presentation which encourages forward reasoning from premises (analytic processing) and when reasoners could give cannot tell responses instead of belief-based responses. Availability of the cannot tell option did not eliminate the believability effect, which was also evident on non-integrable problems. The findings are consistent with a heuristic process that is distinct from analytic processing.

Second, it provided further evidence of individual differences. Premise integration ability predicted analytic processing, over and above the contribution of fluid intelligence, and it was associated with reduced belief bias on invalid problems. Premise integration ability appears to influence the extent of reliance on heuristic versus analytic processing in transitive inference, when premises and conclusions are presented sequentially. Future research will determine whether and how these findings change when premises and conclusions are presented simultaneously.

Third, it advances understanding of conflict detection which is important in dual process models. Conflict detection scores reflecting sensitivity to the belief-logic conflict were unrelated to analytic processing. An intuitive logic (De Neys, 2012) might account for this finding and for the preference for belief-based over cannot tell errors on conflict problems. This process might be sufficient to detect conflict and to identify when a determinate conclusion is appropriate, but not sufficient to resolve the conflict.

References

Andrews, G. (2010). Belief-based and analytic processing in transitive inference depends on premise integration difficulty. *Memory and Cognition, 38,* 928-940.

Andrews, G. & Halford G.S., (1998). Children's ability to make transitive inferences: The importance of premise integration and structural complexity. *Cognitive Development*, *13,* 479-513. DOI:10.1016/S0885-2014(98)90004-1

Andrews, G., & Todd, J. M. (2008). Two Sources of Age-Related Decline in Comprehension of Complex Relative Clause Sentences. In N. B. Johansen (Ed.). *New Research on Short-term Memory*. (pp. 93-123) New York, NY: NovaScience Publishers.

Banks, A. P. (2013) The influence of activation level on belief bias in relational reasoning. *Cognitive Science, 37,* 544-577. DOI: 10.1111/cogs.12017

Bunch, K. M., & Andrews, G. (2012). The development of relational processing in hot and cool tasks. *Developmental Neuropsychology, 37*, 134-152.

Cattell, R. B. & Cattell, A. K. S. (1973). *Culture Fair Test of "g": Scale 2.* Champaign, Ill: Institute for Personality and Ability Testing.

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science, 7*, 28-38.

Evans, J. St. B. T., (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13,* 321-339. DOI: 10.1080/1354678061008825.

Evans, J. St. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: evidence for the dual-process theory of reasoning. *Thinking & Reasoning, 11,* 305-416. DOI: 10.1080/1354678054200005

Maybery, M. T., Bain, J. D., & Halford, G. S. (1986). Information processing demands of transitive inference. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12,* 600-613.

Morley, N. J., Evans, J. St. B., & Handley, S. J. (2004). Belief bias and figural bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology, 57A*, 666-692

Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology, 57A,* 33-60.

Prado, J., Van der Henst, J. B., Noveck, I. A. (2008). Spatial associations in relational reasoning: Evidence for a SNARC-like effect, *The Quarterly Journal of Experimental Psychology, 61*, 1143–1150.

Roberts, M. J., & Sykes, E. D. A. (2003). Belief bias and relational reasoning. *The Quarterly Journal of Experimental Psychology, 56A*, 131-154

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

Stanovitch, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive abilities. *Journal of Personality and Social Psychology, 94*, 672-695.

Stupple, E. J. N., & Ball, L. J. (2008). Belief-logic conflict resolution in syllogistic reasoning: Inspection time evidence for a parallel process model. *Thinking & Reasoning, 14*, 168 – 181.

Stupple, E. J. N., & Ball, L. J., Evans, J. St. B., & Kamal-Smith, E. (2011). When logic and belief collide: Individual differences in reasoning times support a selective processing model. *Journal of Cognitive Psychology, 23*, 931-941.

Thayer, E. S., & Collyer, C. E. (1978). The development of transitive inference: A review of recent approaches. *Psychological Bulletin, 85*, 1327-1343.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., Thomas, C. R., & Miller, B. L. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10*, 119-125.

Table 1

Examples of the Six Problem Types used in Transitive Inference

Inference Type	Premises	Conclusions
Valid Believable (VB)	The elephant is larger than the cow. The dog is smaller than the goat. The mouse is smaller than the dog. The cow is larger than the goat.	Therefore the cow is larger than the dog.
Valid Unbelievable (VU)	The mandarin is sweeter than the lemon. The grapefruit is sweeter than the orange. The mandarin is less sweet than the orange. The grapefruit is less sweet than the cherry.	Therefore the grapefruit is sweeter than the mandarin.
Invalid believable (IB)	The car is faster than the bicycle. The tractor is faster than the train. The car is slower than the train. The tractor is slower than the jet.	Therefore the car is faster than the tractor.
Invalid unbelievable (IU)	The crocodile is more fierce than the shark. The rabbit is less fierce than the zebra. The frog is less fierce than the rabbit. The shark is more fierce than the zebra.	Therefore the rabbit is more fierce than the shark.
Non-integrable believable (NB)	The parrot is bigger than the goldfish. The hamster is smaller than the cat The parrot is smaller than the turtle The dog is bigger than the cat.	Therefore the cat is bigger than the turtle.
Non-integrable unbelievable (NU)	The grass is higher than the flowers. The seedling is higher than the moss. The grass is lower than the shrub. The seedling is lower than the tree.	Therefore the seedling is higher than the shrub.

Table 2.

Zero-order Correlations of Belief, Logic and Interaction Indices[a] with Conflict Detection, Premise Integration and Fluid Intelligence Scores.

	1.	2.	3.	4.	5.	6.
1. Belief index						
2. Logic index	-.22					
3. Interaction index	-.18	-.07				
4. Conflict detection	.24[†]	.14	-.32**			
5. Premise integration	-.13	.47**	-.30*	.22		
6. Fluid intelligence	-.18	.44**	.08	.07	.46**	
Mean	.355	.899	.101	.332	.515	34.540
SD	.487	.617	.420	.388	.299	4.179

[†] $p < .10$; * $p < .05$; ** $p < .01$

[a] Mean values on all indices were greater than zero, smallest t (66) = 1.96, p = .027 (1-tailed)

.

Table 3.

Standard Multiple Regression Analyses Predicting Logic and Interaction Indices.

Criterion	Predictors	*B*	*SE (B)*	ß	sr^2
Logic index	Premise integration	.69	.246	.33	.088**
	Fluid Intelligence	.04	.018	.29	.067*
	Multiple *R* = .534, *F* (2, 64) = 12.78, *p* < .001				
Interaction index	Premise integration	-.35	.165	-.25	.058*
	Conflict detection	-.28	.127	-.26	.066*
	Multiple *R* = .397, *F* (2, 64) = 5.97, *p* = .004.				

p < .05; ** *p* < .01

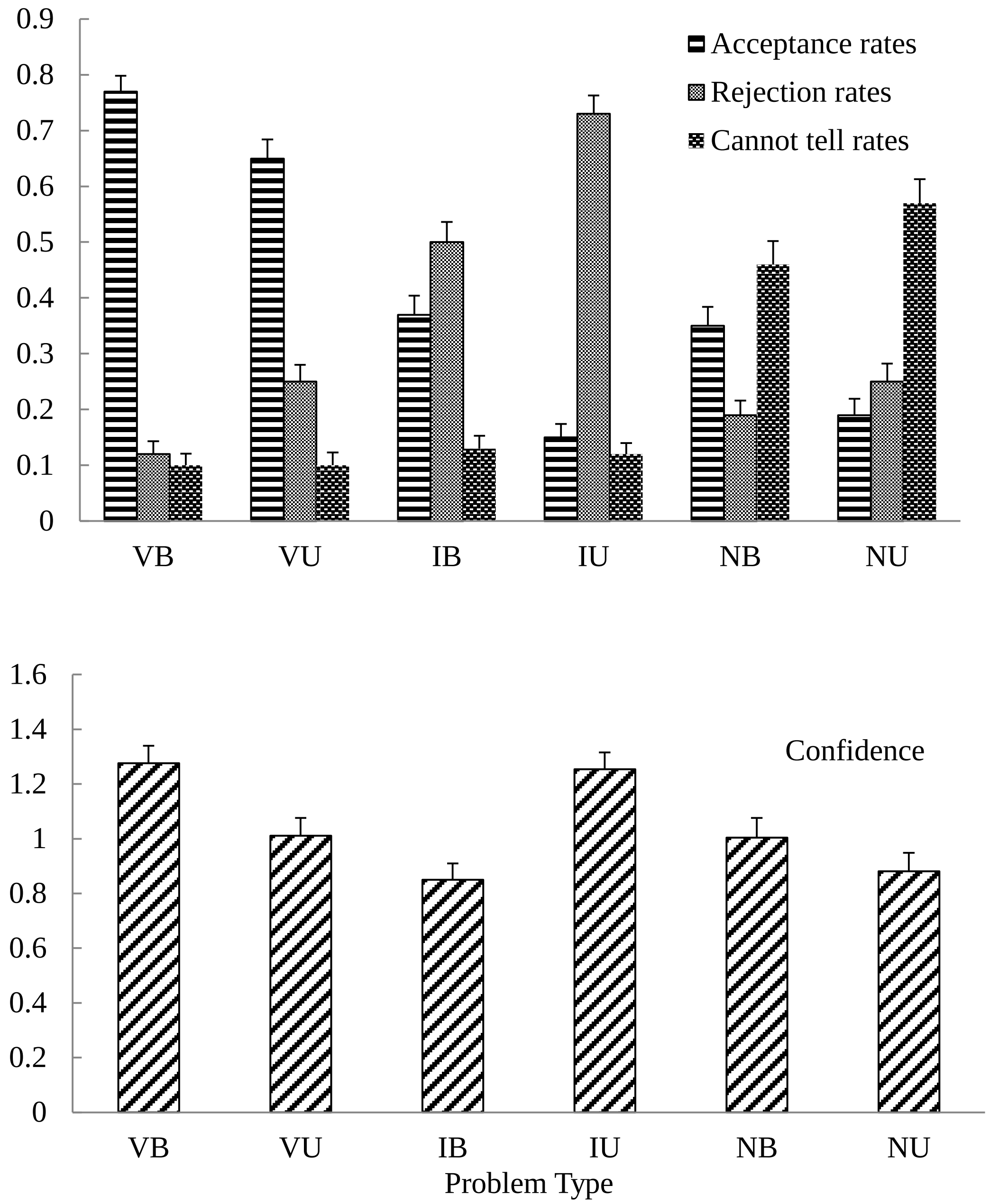

Figure 1. Mean acceptance, rejection, and cannot tell rates (upper) and confidence (lower) for valid believable (VB), valid unbelievable (VU), invalid believable (IB), invalid unbelievable (IU), non-integrable believable (NB) and non-integrable unbelievable (NU) problems. Error bars represent standard errors. Cannot tell rates for all problems were above zero, smallest t (66) = 4.61, $p < .001$.