Investigating single-process and dual-process theories of transitive reasoning: Applying Signal Detection Theory and Signed Difference Analysis.

Nadia Scott

Word Count: 9435

This thesis is submitted in partial fulfilment of the Honours degree of Bachelor of Psychological Science (Honours)

Table of Contents

List of Figures

List of Tables

Abstract

According to influential dual-process theories, reasoning is driven by distinct Type 1 and 2 processing. Type 1 processing is characterised as fast, intuitive and heuristics based, while Type 2 processing is thought to be more effortful, deliberate and requires working memory. However, the dual-process view has faced an increasing amount of criticism over recent years. Single-process theories offer an alternative account, suggesting that reasoning across a range of contexts is reliant on a common assessment of inference strength. The current experiment tested the competing theories using a transitive reasoning task. Key factors relevant to dual-process accounts were manipulated, including premise integration time and working memory demands via premise ordering. Results showed that validity ratings were higher for valid than for invalid arguments, and for believable than for unbelievable conclusions. An interaction between premise ordering (unscrambled vs. scrambled premises) and validity was also observed. These results were consistent with dual-process theories, however, quantitative models were then compared to investigate whether the results were inconsistent with single-process theories. Signal detection theory was applied and dual- and single-process accounts were instantiated as two-dimensional and one-dimensional models, respectively. Model testing via signed difference analysis showed that the observed data do not rule out the simpler single-process, one-dimensional model. This suggests that such single-process models offer a viable account in explaining the underlying cognitive processing intransitive reasoning.

Keywords: Transitive reasoning; dual-process theories; single-process theories; signal detection theory; signed difference analysis; working memory

Declaration

This thesis contains no material which has been accepted for the award of any other degree of diploma in any University, and, to the best of my knowledge, this thesis contains no material previously published except where due reference is made. I give permission for the digital version of this thesis to be made available on the web, via the University of Adelaide's digital thesis repository, the Library Search and through web search engines, unless permission has been granted by the School to restrict access for a period of time.

September 2021

Contribution Statement

Establishing the knowledge gap and formulating the research question was completed through consultation with my supervisor. The ethics approval was provided to me by my supervisor, and the study pre-registration was written in collaboration between myself and my supervisor. My supervisor and I collaborated in designing the study and choosing the key experimental factors. Stimuli examples were provided by my supervisor, which I used to construct a new set of stimuli to be used for this study. I set up the Qualtrics survey and conducted the data analysis for the pilot believability study. The details of the main experiment were decided by my supervisor and I, and the code for the online experiment was written by her Research Assistant. I conducted the literature search and was responsible for participant recruitment and crediting. My supervisor provided me with the R code for data processing and generating graphs, and I produced all tables and diagrams. I ran all the statistical analyses in R and wrote up all aspects of the thesis.

Acknowledgements

I would like to thank my supervisor, Dr. Rachel Stephens, who has provided me with an incredible amount of support throughout this year. Thank you for taking the time to share your expertise and to provide mentorship. I have developed a better understanding of the field and sharpened my knowledge in psychological research and academic writing as a result. Your insightful feedback has helped me to improve my quality of work. Your patience, understanding and warmth has kept me motivated throughout a difficult year. You have been an inspiring and an astounding supervisor and I will forever cherish my experience under your supervision. I would also like to thank Matthew Keasler for helping to implement the online study. Thank you for your attention to detail in writing the code for the experiment and your patience throughout the many revisions of the experiment.

Investigating single-process and dual-process theories of transitive reasoning: Applying Signal Detection Theory and Signed Difference Analysis.

Introduction

Imagine you are going to a meeting at a university that you have never been to before. You are standing at *building a* and you know that *building a* is just to the left of building *b*, but the meeting is located in building *c*. You ask someone for directions, and they tell you that building *b* is to the left of building *c*. From this, you conclude that Building A must be to the left of Building C. How did you come to that conclusion? An important question in the cognitive psychology literature is: what cognitive processes are involved? People's ability to infer the relationships between two objects via the consideration of other known relationships is called *transitive reasoning* (Bouwmeester et al., 2007). Pieces of information or "premises" are given *(a* is to the left of *b, b* is to the left of *c*), and if these premises are true, then you can conclude that the inferred relationship *(a* is to the left of *c*) must also be true. Transitive reasoning is a form of *deductive reasoning* (Lazareva, 2012).

Deductive reasoning refers to the ability to logically infer whether a conclusion is valid (i.e., necessarily follows) from a given set of premises which are assumed to be true (Johnson-Laird, 1999). Deductive reasoning skills are important to understand because they have been shown to predict other cognitive skills like mathematical abilities (Morsanyi et al., 2017) and insight problem-solving (Niu et al., 2007). Deductive reasoning abilities are also considered valuable in the workforce (Carnevale & Smith, 2013; Miulescu et al., 2012; Hunt & Madhyastha, 2010).

Transitive Reasoning

Consider the aforementioned example of transitive reasoning about buildings - that is an example of a locational relationship. A transitive relationship is not necessarily locational, but can be any kind of relationship (e.g., *a* is bigger than *b*; *a* is faster than *b*; *a* is less expensive than *b*). Prado et al. (2008) suggested that the transitive reasoning process involves the encoding of premises, followed by the integration of premises. Premise integration involves mentally assembling the order of premises from top to bottom ($a > b > c$). The conclusion can then be compared with this assembled ordering.

Transitive reasoning abilities have been extensively studied in developmental psychology (see Andrews & Halford, 2002), as popularised through Piaget's Theory of Cognitive Development (Markovits & Barrouillett, 2007). Children begin to show capabilities to reason transitively around the age of five (Phillips et al., 2009). It has also been demonstrated that reasoning abilities show a curvilinear trend, improving with age from childhood to early adulthood and declining again in later life (De Neys & Gelder, 2009; Todd et al., 2019).

Nonetheless, reasoning errors in educated adults still occur in both transitive reasoning and other types of deductive reasoning (Andrews 2010; Bago & De Neys, 2017). Notably, there is a tendency to rely on one's prior beliefs and assumptions rather than the logical structure of arguments when presented with reasoning tasks (see Evans, 2003), also known as the *belief bias effect* (Oakhill & Johnson-Laird, 1985). Conclusions that align with individuals' prior beliefs are more likely to be endorsed than those that are not, regardless of their validity (Evans et al., 1983). This reasoning bias is more prominent when conflict is present between prior beliefs and validity (i.e., when the conclusion is invalid but believable, or valid but unbelievable; Klauer et al., 2000). A core research question in the deductive reasoning literature is to understand the cognitive processes that underlie reasoning and belief-bias.

A task commonly used to study belief-bias in deductive reasoning is called the *argument evaluation task*, in which participants are presented with a set of premises and a conclusion and are instructed to determine whether the conclusion logically follows from the premises (e.g., Andrews, 2010; Stephens et al., 2020). In this procedure, four argument types are generally presented: valid-believable, invalid-believable, valid-unbelievable and invalid-unbelievable which are shown in Table 1. Invalid-believable and valid-unbelievable are known as conflict problems because the logical validity and believability of arguments are in conflict in terms of the subjective strength of an argument (see Table 1). The main findings often yielded from argument evaluation tasks are that: a) endorsement rates (i.e., endorsing an argument as deductively valid) are higher for valid than invalid problems, but b) endorsement rates are also higher for believable than unbelievable problems, and c) a greater believability effect is sometimes observed for invalid than valid problems (Andrews, 2010). The latter two effects indicate belief bias.

Table 1

Examples of Transitive Reasoning Problem Types

Argument and Conclusion Type	**Example Premises and Conclusion**	**Conflict Type**
Valid-believable	Elephants are bigger than dogs Dogs are bigger than mice Therefore, elephants are bigger than mice	No-conflict problem
Valid-unbelievable	Mice are bigger than dogs Dogs are bigger than elephants Therefore, mice are bigger than elephants	Conflict problem
Invalid-believable	Mice are bigger than dogs Dogs are bigger than elephants Therefore, elephants are bigger than mice	Conflict problem
Invalid-unbelievable	Elephants are bigger than dogs Dogs are bigger than mice Therefore, mice are bigger than elephants	No-conflict problem

Dual-Process Theories

Dual-process theories have been used to provide an explanation for the belief bias effect (Andrews, 2010; Andrews & Michelic, 2014). Dual-process theories have been highly influential, and have been applied to many settings, such as: explaining the results of neurological research (e.g., Prado et al., 2011; Goel et al., 2017), understanding people's healthiness judgements of nutrition labels (Sanjari et al., 2017), guiding behavioural nudging (Blom et al., 2021), and informing medical diagnosis and educational approaches (see Stephens et al., 2020). Under a dual-process view, it is proposed that reasoning is driven by two different types of cognitive processes, referred to as Type 1 and Type 2 processing. (Evans, 2003; Evans, 2010; Evans & Stanovich, 2013). Type 1 processing is characterised as generally being fast, intuitive, and heuristic; it is believed to be more reliant on individuals' prior beliefs and perceptions, therefore, Type 1 processing tends to invoke more biased responses. Meanwhile, Type 2 processing is characterised as being more effortful, more analytical, and requires working memory. Type 2 processing tends to be slower to activate but can produce more accurate responses.

There are multiple of variants of dual-process theories, with different ideas about how Type 1 and Type 2 processing interacts. Researchers such as Evans (2007) have considered different models of how the two reasoning processes interact, including pre-emptive, parallel-competitive and default interventionist models. For example, the well-known default-interventionist model suggests that the heuristic, Type 1 processing typically occurs from the onset of argument evaluation, subsequently followed by activation of the analytical, Type 2 processing (Evans, 2007). There are also other conceptions of when and how Type 1 and 2 processing are activated during reasoning, such as the parallel processing model by Sloman (1996) that proposes that both Type 1 and Type 2 processing occur in parallel. Nonetheless, the variants of dual-process theory all revolve around the notion that there are two distinct processes of reasoning which compete to guide responses in tasks like the argument evaluation task. The two processes generally work cooperatively, however, during conflict resolution, it is believed that the domination of Type 1 processing typically gives rise to the belief-bias effect (Andrews, 2010).

Beyond belief bias, critical experimental evidence that has been used to support dual-process theories is based on task dissociations (Stephens et al., 2018). Task dissociations have been demonstrated through experimental factors that were designed to target one type of processing without impacting the other (Evans & Stanovich, 2013). Key experiment factors have been shown to increase belief bias and reduce sensitivity to validity, purportedly through the suppression of Type 2 processing; namely, factors including response deadlines (e.g., Evans & Curtis-Holmes, 2005) and working memory load (e.g., De Neys, 2006; Howarth et al., 2016). Evans and Curtis-Holmes (2005) found an interaction between time and believability. Particularly, the believability effect was greater in the condition where premise encoding time and conclusion evaluation time was shorter. The results suggested that participants were more affected by the believability of conclusions when they had less time for argument evaluation compared to those who were in a free-time condition. The authors concluded that this was consistent with the belief-first dual process models (e.g., default interventionist) and that having time constraints limited the activation of Type 2 processing, thus resulting in more reliance on Type 1 processing. Additionally, to demonstrate dual-processes, measurement or manipulation of working memory capacity is often performed in experiments as the involvement of working memory is presumed to be a key feature of Type 2 processing (Evans, 2010). Experiments have shown that reasoners are affected by a concurrent task that imposes a higher working memory load, resulting in more belief-based than validity-based responses when evaluating arguments (e.g., De Neys, 2006). Moreover, individuals with higher fluid intelligence (and presumably, higher working memory capacity) are generally less affected by belief bias (Conway et al., 2003; Stanovitch & West, 2008).

Problematically, although dissociation evidence may be consistent with dual-process theory, recent research has highlighted that such evidence is unconvincing and does not *compel* the existence of multiple underlying processes (e.g., Stephens et al., 2018). One issue is that establishing a "pure" dissociation (which includes no measurable effect in one task condition, with a substantial effect in another task condition) relies on the acceptance of the null hypothesis, which is problematic in itself (see Hayes et al., 2018; Stephens et al., 2018). Regardless, the observed dissociations have not been that clear-cut for the majority of effects that have been observed in previous studies. Rather than pure dissociations, results of experimental manipulations (i.e., time and working memory) in previous research usually revealed weaker evidence, such as a small versus medium or large effect of certain variables (e.g., a smaller effect of validity under time pressure than under no pressure). However, even if pure dissociations were observed, it has been shown that they are still not necessarily inconsistent with single-process theories (see Newell & Dunn, 2008). Task dissociations have been re-examined in other areas such as recognition memory (Dunn & Kirsner, 1988), memory development (Hayes et al., 2016) and category learning (Stephens et al., 2019a), to test for compelling evidence for multiple underlying cognitive processes. It has been demonstrated that a single underlying latent variable can provide an alternative explanation for task dissociations observed in these areas, thus, showing dissociations to be insufficient as evidence for more than one underlying process (Hayes et al., 2018).

Another problem for dual-process theory is that an increasing number of studies have been published in recent years that are inconsistent with the traditional dual-process accounts. These findings have led theorists to develop newer, more complex versions of dual-process models (e.g., Bago & De Neys, 2017; Handley & Trippas, 2015) which combine some of the key features of the traditional sequential models (i.e., default interventionist) and parallel models. For example, Bago and De Neys' (2017) revised dual-process model proposes that Type 1 processing considers both "heuristic based" (believability) AND "analytical" (validity) argument cues, rather than the latter being considered predominantly via Type 2 processing. The model then proposes that reasoning performance is determined by the strength of different types of "intuition". Such models have further blurred the distinction between Type 1 and Type 2 processes, which begs the question of whether the distinction remains theoretically useful. Authors such as Keren and Schul (2009) have discussed the questionable nature of dual-process theories in the past and queried whether such theories can truly provide scientific advancement. Thus, an increasing amount of research has been conducted to determine whether a viable alternative account of reasoning can provide new insight to the underlying cognitive processes that drive reasoning (e.g., Hayes et al., 2018; Stephens et al., 2020).

Single-Process Theories

As an alternative account, single-process theories of reasoning assume that arguments are assessed based on a common underlying cognitive process, regardless of whether reasoners are making fast versus slow, or seemingly intuitive versus deliberate judgements (Hayes et al., 2018; Kruglanski & Gigernzer, 2011). There are several variants of single-process theories with different ideas of what the common process might be. For example, reasoners may utilize Bayesian probability estimation and belief revision while performing argument evaluation tasks (Oaksford & Chater, 2001). Others have also proposed an important account based on a signal detection framework, which proposes that reasoning involves a subjective assessment of argument "strength" and a criterion threshold for endorsing an argument as valid (Hayes et al., 2018; Rips, 2001; Rotello & Heit, 2009).

Single-process models based on the signal detection framework have been shown to account for a wide range of reasoning performance across both new studies and re-analysed databases, including judgements under time constraints (fast vs. slow) or under working memory load (Stephens et al., 2018, 2020). These models focussed on comparing deduction judgements (is the conclusion valid?) with induction judgements (is the conclusion plausible?) under time pressure, working memory load, and so on, under the assumption that deduction reflects more Type 2 processing, while induction reflects more Type 1 processing (e.g., see Evans et al., 2010). Under the signal detection framework, belief-bias effects and dissociations between induction and deduction judgements may be the result of differences in response threshold, with no need to posit distinct argument assessments based on Type 1 or 2 processing (Stephens et al., 2018).

Previous successful single-process signal detection models have been tested against experiments that compare induction and deduction judgements in an attempt to capture more Type 1 versus Type 2 processing, respectively (under the competing dual-process account; e.g., Hayes et al., 2018; Stephens et al., 2018). However, it is possible that the dual-process account is correct but these single-process models have not been ruled out because induction and deduction judgements do not sufficiently differentiate between Type 1 and Type 2 processing. Perhaps, to reveal the two distinct processing types, alternate dependent variables or tasks are needed to make the processing types more distinguishable (Stephens et al., 2020). Accordingly, the goal of the current experiment is to apply the signal detection approach, however, with alternative dependent variables in an argument evaluation task. Instead of using induction and deduction judgements to try to capture any distinction between Type 1 and Type 2 processing, a focus on working memory demands (low vs. high) during deduction judgements would be a more rigorous test of the theories – especially given that the involvement of working memory is a defining feature of Type 2 processing (see Evans & Stanovich, 2013).

Working Memory and Transitive Reasoning

Working memory capacity can be defined as a central and singular source of limited capacity, measured by the number of items held active in short-term memory storage (Evans, 2010). There is a strong link between working memory and transitive reasoning abilities (Evans 2008; Halford et al., 2007). In transitive reasoning, the capacity limit is quantified as the number of interrelationships between elements that can be kept active in working memory (Halford et al., 2007).

In transitive argument evaluation tasks, participants are typically given premises about relations between a set of elements, and then are asked to evaluate two non-adjacent elements in the conclusion (see Table 1). Encoding of transitive premises requires integrating the elements and arranging them in an ordered array (Andrews, 2010). The integration of premises in transitive reasoning is a complex "ternary" process due to the need for two binary-relational premises (e.g., elephants > dogs; dogs > mice) to be integrated in order to construct an order (elephants > dogs > mice; Halford et al., 1998). The binary-relationships between elements are stored in memory "chunks" (e.g., elephants > dogs = chunk 1, dogs > mice = chunk 2). These memory chunks assist in reducing the cognitive load required by allowing items to be grouped together as one piece of information to be kept in working memory. The working memory of adults generally has a capacity of approximately 3-5 chunks that can be kept active in working memory, while children and the elderly can retain fewer (Halford et al., 2007). To integrate the elements into an order, the premises chunks are mapped onto an ordering schema to assist in inferring the order of the elements in the conclusion. For example, a mental model might be built based on size position from biggest to smallest in this instance (see Table 2). However, a processing load is imposed during this process as each premise needs to be considered to bind each element to a position (see Table 2; Halford et al., 2007). Thus, premise encoding and integration components in transitive reasoning are thought to be more cognitively demanding than conclusion evaluation (Andrews, 2010).

Table 2

An example of relational premises integrated to generate an ordered mental model (based on Halford et al., 2007)

Mental Model	Premise Element
Biggest	Elephants
Middle	Dogs
Shortest	Mice

Note. Mental model and premises for the argument, "Elephants are bigger than dogs. Dogs are bigger than mice. Therefore, elephants are bigger than mice."

Transitive inference problems allow for manipulations that increase the difficulty of the task to impose a higher cognitive load. The premise integration difficulty can be manipulated by increasing the number of premises and scrambling their presentation order. In a five-element series – $a > b$, $b > c$, $c > d$, $d > e$, premises presented in this sequential order impose less working memory load as only one premise needs to be considered at a time, allowing each element to be added to the chain as it is encountered to form the $a > b > c > d > e$ order (Andrews & Halford, 1998). In contrast, in a scrambled presentation order (e.g., $d > e$, $b < a$, $b > c$, $c > d$; see Table 3), each premise must be held active in working memory until all the necessary pieces of information are acquired to construct an overall order (Andrews, 2010).

Table 3

Unscrambled and Scrambled Examples of Five-Term Transitive Inferences

Order	Premises	Elements
Unscrambled	Elephants are bigger than hippos	$a > b$
	Hippos are bigger than cows	$b > c$
	Cows are bigger than dogs	$c > d$
	Dogs are bigger than mice	$d > e$
Scrambled	Dogs are bigger than mice	$d > e$
	Hippos are smaller than elephants	$b < a$
	Hippos are bigger than cows	$b > c$
	Cows are bigger than dogs	$c > d$

Note. *a* = elephants, *b* = hippos, *c* = cows, *d* = dogs, *e* = mice

Premise presentation order has been shown to affect transitive reasoning performance. In an argument evaluation experiment by Andrews (2010), greater believability effects were observed when premises were presented in a scrambled order compared to unscrambled. This believability effect indicated that participants were more susceptible to belief-bias when working memory load was increased due to premises being more difficult to integrate. In contrast, a greater validity effect was observed when premises were unscrambled compared to scrambled. This validity effect suggests that people's ability to differentiate between valid and invalid arguments is compromised when premise integration is difficult. Drawing on these findings from Andrews (2010), manipulation of premise presentation order will be a critical factor to try to differentiate between Type 1 versus Type 2 processing under a dual-process account.

The Impact of Time Pressure on Transitive Reasoning

In order to test single-process signal detection models, additional factors are needed to further impact reasoning when premises are scrambled versus unscrambled. Time pressure is important to include, as it has been a key factor in supporting dual-process accounts (Evans, 2010). Through several argument evaluation experiments, Andrews (2010) and Andrews and Michelic (2014) investigated the effect of premise presentation time in transitive reasoning. Experiments were conducted to measure acceptance rates (i.e., how often participants accept conclusions as deductively valid) of transitive arguments under different premise integration time pressure: faster presentation (10 seconds) or slower presentation (20 seconds). Overall, it was concluded that premise integration time influenced whether Type 1 or Type 2 processing was more likely to take place. The validity effect was greater in the slower presentation condition compared to the faster presentation condition, thus, suggesting that validity was processed more thoroughly when more time was available. In turn, the believability effect was greater in the faster presentation condition than in the slower condition, indicating that participants were more prone to belief-bias when less time is available.

The studies by Andrews (2010) and Andrews and Michelic (2014) were conducted and interpreted with the assumption that dual-process accounts are true and the aim was to differentiate between different versions of dual-process accounts. It was concluded that the results were consistent with both the default-interventionist and parallel-process dual-process models. However, the authors also acknowledged that the results may not be uniquely predicted by such dual-process models and may not be inconsistent with single-process theories. Banks and Hope (2014) similarly conducted an experiment investigating belief-bias in transitive reasoning, with the aim to differentiate between dual-process accounts without considering single-process theories. It is therefore important that single-process accounts of transitive reasoning are explicitly tested against dual-process accounts.

The Current Study

The current study sought to build on Andrews' (2010) experiments on transitive reasoning. Following this previous work, I manipulated premise integration time and working memory demand via premise presentation order, to confirm whether transitive reasoning performance is differentially affected by these factors. However, I combined the manipulations of premise integration time and working memory demand into a single large experiment, to more rigorously test competing single-process and dual-process theories. The theories were tested via novel variants of single and dual-process signal detection models developed by Stephens et al. (2018). Due to the complexity of the models, multi-factor experiments are needed to help distinguish between them.

The first aim was to test whether there are differential effects of validity and believability for both fast versus slow premise presentation time, and scrambled versus unscrambled premises. Premise scrambling should increase working memory demand and thus decrease the effect of validity and increase the effect of believability. Such effects may be similarly produced and further exacerbated by reducing premise presentation time. The second aim was to then test whether a key single-process signal detection model could be ruled out in favour of a rival dual-process signal detection model. The models will be tested in their most general form using a novel technique called Signed Difference Analysis (Dunn & James, 2003), described below.

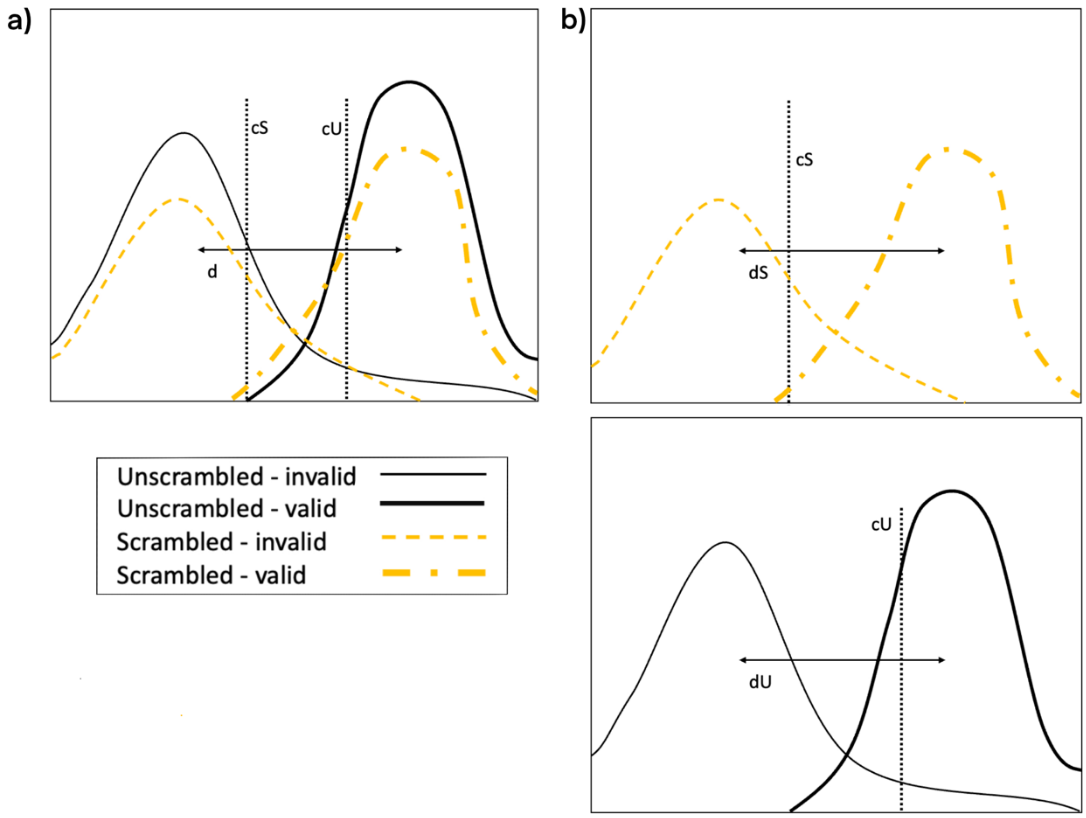
Signal Detection Theory

Signal detection theory has been applied to many diverse areas of cognitive psychology such as speech perception, memory, and eyewitness identification (Pastore & Scheirer, 1974; Wixted & Mickes, 2014). Using signal detection theory in reasoning has proven advantageous as it is applicable to a wide range of argument evaluation tasks with varying argument complexity and structures (Stephens et al., 2018). According to Stephens et al. (2018, 2020), the evidence relating to a discrimination judgement (e.g., valid vs. invalid) is distinguished from the criterion that the evidence is being compared against. It is assumed that subjective argument strength is assessed on continuous dimension(s): one dimension (1D) to instantiate a single-process account or two dimensions (2D) to instantiate a dual-process account. The most important point is that if there are two distinct cognitive processes that underlie reasoning, distinct assessments of argument strength should be observed when working memory demands are high versus low, with each based more on the response from Type 1 or Type 2 processing, respectively. Contrastingly, if reasoning relies on a common cognitive process, then there will be only one type of argument assessment across different levels of working memory demand. These model types have previously been applied to induction and deduction judgements (e.g., Stephens et al., 2018, 2020), but the current work adapts them to account for reasoning judgements when working memory demands are high versus low (scrambled vs. unscrambled premises).

For each signal detection model, there are distinct distributions of argument strength in the 1D or 2D space for valid and invalid arguments, for both unscrambled and scrambled conditions. The extent to which participants distinguish between the valid and invalid arguments is reflected by the distance between the distributions (see Figure 1). For the 2D model, two discriminability parameters (*dU* for the unscrambled condition and *dS* for the scrambled condition) are included, while the 1D model has only one single discriminability parameter (*d*). Decision thresholds are assumed to be set by participants during the argument evaluation task. Thus, only arguments that sit above the criterion in subjective strength will be endorsed. According to the models in Figure 1, there are distinct criteria for unscrambled and scrambled conditions (*cU* vs. *cS*). Simpler models are also possible whereby restrictions are placed on the criteria parameters (see Stephens et al., 2018), but the current project focuses on these two most general single- and dual-process models. Furthermore, because the decision thresholds for scrambled and unscrambled conditions are independent, the 1D and 2D model variants are referred to as the *independent-1D* and *independent-2D models*, respectively (Stephens et al., 2018).

Figure 1

One- and Two-Dimensional Signal Detection Models in an Argument Evaluation Task



Note. Signal detection models of argument evaluation tasks under high versus low working memory demand; scrambled versus unscrambled premises, respectively. Low working memory demand refers to the "unscrambled" condition where less working memory is required to integrate transitive premises. High working memory demand refers to the "scrambled" condition. a) three parameter single-process, independent-1D model with a discriminability parameter, *d*, and decision criteria (thresholds) for unscrambled *c*U and scrambled, *cS*. b) four-parameter dual-process, independent-2D model with separate discriminability parameters for unscrambled, *dU,* and scrambled, *dS*, conditions, but the same criteria parameters as the 1D model.

Signed Difference Analyses

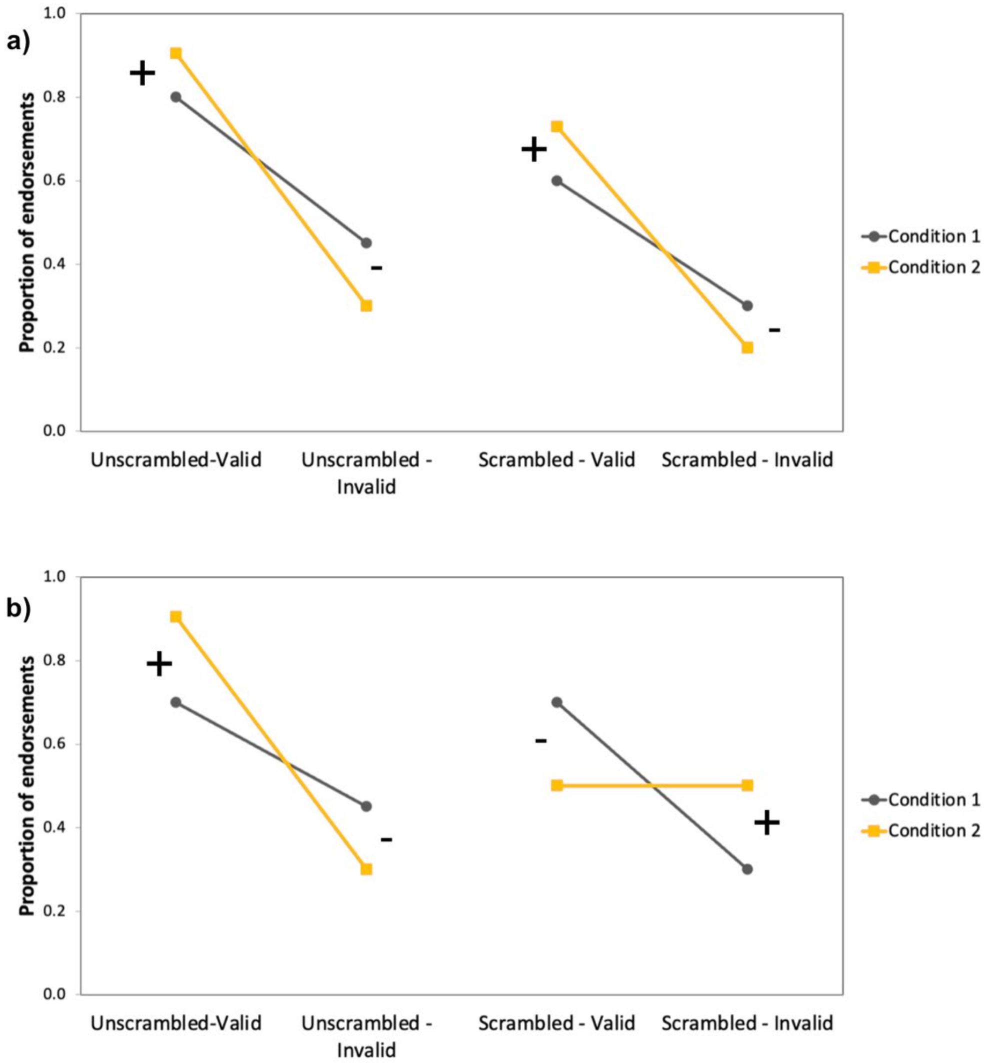
Stephens et al. (2018, 2020) demonstrated how signed difference analysis (SDA; Dunn & James, 2003) can be used to test the competing independent-1D and -2D signal detection models in their most general form. As the true form of the distributions of argument strength is unknown, SDA facilitates the test of rival signal detection models by making only minimal assumptions. SDA does not assume Gaussian distributions of subjective strength, as per standard signal detection approaches, but simply assumes a monotonic relationship between outcome variables and combinations of model parameters. In other words, if the model parameters shift positively as a result of an experimental manipulation, then an increase or no change (but no decrease) in argument endorsements should be observed. A key advantage of SDA is that if a model is ruled out by the data, then this result cannot be attributed to a misspecification of strength distributions. Models are tested based on their core features – the model parameters – rather than on their more auxiliary assumptions about distributional form (for further discussion see Stephens et al., 2018).

The current data are grounded in four dimensions in this application of SDA, based on endorsement rates of valid and invalid arguments for ordered or scrambled premises, as shown in Figure 2 along the x-axis. Endorsement rates for ordered-valid, ordered-invalid, scrambled-valid, and scrambled-invalid conditions constitute the “dependent variables”, and SDA involves testing for ordinal patterns of differences between conditions across the set of dependent variables. The different conditions in this study are based on factorial combinations of premise presentation time (fast vs. slow) and believability (low vs. high).

Stephens et al. (2018) showed that various 1D and 2D signal detection models have a number of “permitted” and “forbidden” ordinal data patterns. Hypothetical examples of two of these data patterns are shown in Figure 2. Figure 2a illustrates a data pattern consistent with improved discrimination between valid and invalid conclusions in Condition 2 compared to Condition 1 for both unscrambled and scrambled premises. The observed differences can be captured by signed difference vectors, which in this instance is (+ - + -). To elaborate, for validity discrimination in unscrambled-valid, there is an increase in endorsements in Condition 2 compared to Condition 1, therefore the difference between Condition 1 to Condition 2 is positive, hence the assigned (+) signed difference in the vector. Whereas, for unscrambled-invalid conclusions, there is a decrease in endorsements in Condition 2 compared to Condition 1, hence the assigned (-) signed difference in the vector. There are multiple data patterns permitted by both the independent-1D and -2D models, however only one data pattern that is forbidden by the 1D model which is shown in Figure 2b. The (+ - - +) signed difference vector (the reversed cross-over patterns) is forbidden because it suggests opposing shifts in validity discrimination between unscrambled and scrambled conditions. Due to having one dimension of argument strength across the two conditions, it is impossible for validity discrimination performance to be better for unscrambled premises and simultaneously worse for scrambled premises (or vice versa). Therefore, if the forbidden pattern is observed, it can be concluded that the 1D model is ruled out in favour of the 2D model. Accordingly, the current study aimed to examine whether the forbidden data pattern would be observed, as a rigorous test of the competing models.

Figure 2

Hypothetical Data Patterns in Signed Difference Analysis



Note. V = Valid, IV = Invalid. The conditions could be based on a combination of high/low believability and premise presentation time (10 vs. 20 seconds). a) an example of the ordinal data pattern permitted by both independent-1D and -2D models corresponding to the signed difference vector (+ - + -) b) an example of the ordinal data pattern forbidden by the independent-1D model which corresponds to the sign difference vector (+ - - +). A signed difference vector refers to the vector of observed differences between two conditions.

Method

Pilot Believability Study

As the believability of conclusion statements was manipulated in the main study, a pilot study to confirm the difference in believability was first conducted. In an online survey, participants were shown 36 statements in randomised order. The 36 statements were presented in both believable and unbelievable forms, with random allocation to participants. Believable forms of statements were those that should have been consistent with basic general knowledge (e.g., dogs are bigger than mice), while unbelievable forms of statements were generally inconsistent with such knowledge (e.g., mice are bigger than dogs). Participants were asked to rate the believability of each statement on 1-5 scale (extremely believable, somewhat believable, neither believable or unbelievable, somewhat unbelievable, and extremely unbelievable).

A total of 58 participants were surveyed (9 males, 49 females, Mean age = 19.95, *SD* = 5.28). Participants were first-year Psychology students and they received course credit. They did not participate in the main experiment. Based on the mean ratings, the 32 items that showed the biggest difference in believability between the two forms were selected for the main study (see Appendix A for the full list). There was a significant difference between believable and unbelievable conclusion statements $t(57) = 32.38$, $p < .001$. Believable statements had a mean rating of 1.43 (*SD* = 1.43), compared to a mean rating of 4.42 for unbelievable statements (*SD* = 0.41).

Design

The main argument evaluation experiment used a 2 (validity: valid, invalid conclusions) x 2 (believability: believable, unbelievable conclusions) x 2 (premise order presentation: unscrambled, scrambled presentation) x 2 (premise presentation time: fast, slow presentation) mixed design. Validity, believability, and premise order were manipulated within participants. Premise presentation time was manipulated between participants. The dependent variable was ratings of confidence in the validity of the conclusions. The study methodology was pre-registered (see Appendix B).

Participants

Participants were 122 first-year Psychology students at the University of Adelaide and received course credit for participation. Out of the total sample, 65 participants were randomly allocated to the fast condition (53%) and 57 were allocated to the slow condition (47%). The eligibility criteria for participation were being over the age of 18 or a student at the University of Adelaide, and being fluent at English. Mean age was 20.39 (SD = 4.07), and 35 participants identified as males (28.7%), 85 identified as females (69.7%), and 2 identified as non-binary/other (1.6%). Note that the intended number of participants as stated on the study pre-registration (see Appendix B) was 100. However, due to a clerical error, more participants signed up for the study than anticipated[1].

Stimuli

Similar to Andrews (2010), the argument evaluation task was made up of four problem types by crossing validity and believability. Each transitive argument contained five elements, *a-e*. The elements together made up four premises that when integrated, constructed the order $a > b, b > c, c > d, d > e$, as shown in Table 4. Valid conclusions were always $b > d$ and invalid conclusions were $d < b$. Five elements were used and the conclusion included only intermediate elements *b* and *d* to prevent a simple "labelling" strategy (Andrews & Halford, 1988). In other words, participants would be less likely to simplify the task and associate a particular element with a label according to where it appeared in the premises. For example, in an easier three-term transitive inference problem ($a > b, b > c = a > c$), a labelling strategy can be employed by labelling *a* as "large" as it appears once as the large element and *c* as "small" as it appears once as the small element (Andrews & Halford,

[1] The data analyses were also conducted based on only the first 100 participants, as per the study pre-registration. However, none of the key results were affected.

1988). One potential methodological limitation that was noted in Andrews' (2010) study was that some problems had some unbelievable premises, thus, potentially confounding premise content with conclusion validity and believability. This was addressed in the current study by implementing a nonsense term (e.g., *zoots*; see Table 4), similar to Banks and Hope (2014). The nonsense term was always applied to element *c* and did not appear in the conclusions.

The argument evaluation task included four argument types: valid-believable, valid-unbelievable, invalid-believable, invalid-unbelievable. In allocating the content to elements *a-e* for the four argument types, adjustments had to be made to accommodate the believability and validity of the conclusion statements. Thus, for valid-unbelievable and invalid-believable arguments, *b* and *d* elements were switched. For example, if valid-believable and invalid-unbelievable arguments had *b* = dogs and *d* = mice; *b* then became mice and *d* became dogs in valid-unbelievable and invalid-unbelievable arguments (see Table 3).

There were 32 sets of argument content in total, based on the pilot study. For each set of content, eight different possible argument conditions were created (premise order × validity × believability). Out of the eight conditions, four were "unscrambled", while the other four were "scrambled". For the unscrambled condition, premises and elements were always presented sequentially from *a* to *e* and organised as seen in Table 4. Meanwhile, the scrambled condition had four different fixed orders which were randomly allocated across the whole stimuli set (See Appendix C). The elements in the scrambled condition were scrambled both within each premise (e.g., *b* < a instead of a > *b*) and across premises. The four fixed orders were randomly chosen from the stimuli set used in Andrews' (2010) study. Each participant saw only one of the eight argument conditions for each of the 32 contents, and saw each argument content once.

Table 4

Examples of Eight Types of Argument Created for a Set of Content

Argument condition	Example premises and conclusion	Elements
Unscrambled-Valid-Believable	Elephants are bigger than dogs	$a > b$
	Dogs are bigger than zoots	$b > c$
	Zoots are bigger than mice	$c > d$
	Mice are bigger than worms	$d > e$
	Therefore, dogs are bigger than mice	$b > d$
Unscrambled-Valid-Unbelievable	Elephants are bigger than mice	$a > b$
	Mice are bigger than zoots	$b > c$
	Zoots are bigger than dogs	$c > d$
	Dogs are bigger than worms	$d > e$
	Therefore, mice are bigger than dogs	$b > d$
Unscrambled-Invalid-Believable	Elephants are bigger than mice	$a > b$
	Mice are bigger than zoots	$b > c$
	Zoots are bigger than dogs	$c > d$
	Dogs are bigger than worms	$d > e$
	Therefore, dogs are bigger than mice	$d > b$
Unscrambled-Invalid-Unbelievable	Elephants are bigger than dogs	$a > b$
	Dogs are bigger than zoots	$b > c$
	Zoots are bigger than mice	$c > d$
	Mice are bigger than worms	$d > e$
	Therefore, mice are bigger than dogs	$d > b$
Scrambled-Valid-Believable	Elephants are bigger than dogs	$a > b$
	Mice are smaller than zoots	$d < c$
	Worms are smaller than mice	$e < d$
	Dogs are bigger than zoots	$b > c$
	Therefore, dogs are bigger than mice	$b > d$
Scrambled-Valid - Unbelievable	Elephants are bigger than mice	$a > b$
	Dogs are smaller than zoots	$d < c$
	Worms are smaller than dogs	$e < d$
	Mice are bigger than zoots	$b > c$
	Therefore, mice are bigger than dogs	$b > d$

	Elephants are bigger than mice	$a > b$
	Dogs are smaller than zoots	$d < c$
Scrambled-Invalid Believable	Worms are smaller than dogs	$e < d$
	Mice are bigger than zoots	$b > c$
	Therefore, dogs are bigger than mice	$d > b$
	Elephants are bigger than dogs	$a > b$
	Mice are smaller than zoots	$d < c$
Scrambled-Invalid-Unbelievable	Worms are smaller than mice	$e < d$
	Dogs are bigger than zoots	$b > c$
	Therefore, mice are bigger than dogs	$d > b$

Note. For valid-believable and invalid-unbelievable arguments, the elements were: *a* = elephants, *b* = dogs, *c* = zoots, *d* = mice, *e* = worms. For invalid-believable and valid-unbelievable arguments, the elements were *a* = elephants, *b* = mice, *c* = zoots, *d* = dogs, *e* = worms.

Procedure

The main experiment was conducted online using custom code. Participants were randomly allocated to either the fast (20 seconds premise presentation time) or the slow (10 seconds premise presentation time) condition and participated individually. Participants were given instructions to assess the logical validity of the reasoning problems. They were instructed to always assume that the premises are true, even if they include a nonsense word. Clear definitions of valid and invalid conclusions were given. Valid conclusions were defined as those that necessarily follow from the premises, while invalid conclusions were defined as those that do not necessarily follow from the premises. Participants were also informed of the time they had to read the premises, according to their assigned condition. Instructions were given for participants to read the arguments carefully and an example transitive reasoning argument was shown. Participants were given three practice trials before beginning the main experiment and then were told that there would be 32 problems in the main task. For each participant, the 32 trails were shown in random order. Each participant saw four instances of each argument condition (premise order × validity × believability). The content was randomly allocated to the argument conditions. Note that this is a methodological improvement to Andrews' (2010) studies, which used fixed allocation.

For each trial, an appropriate "begin the experiment" or "see next trial" button appeared for participants to indicate when they were ready to begin the next trail. After a brief interval of 0.5 seconds, participants began by reading the premises which all appeared at the same time. The premises were displayed for either 10 or 20 seconds, then the conclusion appeared underneath. Participants were allowed 5 seconds to read the conclusion with the premises still being displayed. After 5 seconds, the premises and conclusion disappeared, and the response buttons were presented. Participants were asked to rate their confidence in the validity of the conclusion, with response options: definitely invalid, probably invalid, possibly invalid, possibly valid, probably valid, and definitely valid. These responses were coded as 1-6. There was no limit on response time. No feedback was provided on reasoning problems.

Results

Analysis of Ratings

The mean ratings for all 16 conditions are summarised in Table 5 and presented in Figure 3. A mixed factorial 2 (validity) × 2 (believability) × 2 (premise presentation order) × 2 (premise presentation time) ANOVA revealed significant main effects of validity, $F(1,120) = 135.09$, $p < .001$, $\eta^2 = 0.272$, and believability, $F(1,120) = 85.62$, p < .001, $\eta^2 = 0.169$. Mean ratings were higher for valid arguments ($M = 4.26$, $SD = 1.32$) than invalid arguments ($M = 2.85$, $SD = 1.3$), and for believable arguments ($M = 4.09$, $SD = 1.37$) than unbelievable arguments ($M = 3.03$, $SD = 1.41$). Crucially, as predicted by dual-process theories, a significant interaction between premise order and validity was observed, $F(1, 120) = 56.99$, p

< .001 , $\eta^2 = 0.05$. There was higher validity discrimination when working memory demands were lower (i.e., in the unscrambled condition; see Table 5). However, unexpectedly, there were no effects of premise presentation time, and all other effects were non-significant (see Table 6)

Due to the non-significant effects of premise presentation time, participants' response times were examined. As the response deadline itself was not controlled, it was possible that participants in the fast (10 seconds) condition were compensating by taking longer to reflect on the premises and conclusion before responding. However, there was no evidence of such compensatory behaviour. Response times were similar across the two groups; if anything, response times for participants in the "fast" condition (*M* = 6.09 seconds, *SD* = 118.58) were slightly shorter than participants in the "slow" condition (*M* = 6.26 seconds, *SD* = 105.88).

Table 5

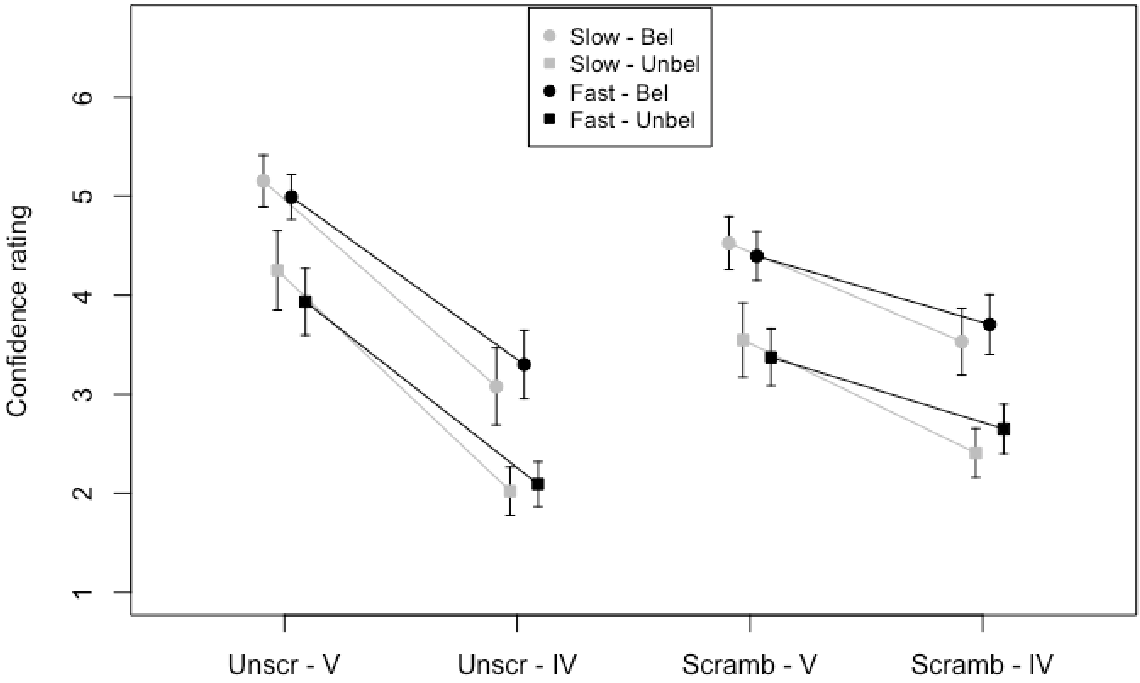
Ratings of Confidence in the Validity of Conclusions

	Long Premise Presentation Time (20 seconds)				Short Premise Presentation Time (10 seconds)			
	Unscrambled Condition		Scrambled Condition		Unscrambled Condition		Scrambled Condition	
Argument Type	M	SD	M	SD	M	SD	M	SD
VB	5.15	0.98	4.53	1.01	4.99	0.91	4.40	0.99
VU	4.25	1.52	3.55	1.41	3.94	1.36	3.37	1.16
IB	3.08	1.47	3.53	1.26	3.3	1.39	3.70	1.21
IU	2.02	0.93	2.41	0.93	2.09	0.91	2.65	1.01

Note. VB = valid-believable, VU = valid-unbelievable, IB = invalid-believable, IU = invalid-unbelievable.

Figure 3

Mean Confidence in the Validity of the Conclusions Ratings



Note. Slow = 20 seconds premise presentation time, Fast = 10 seconds premise presentation time, V = Valid conclusions, IV = Invalid conclusions. Unscr = Unscrambled and Scramb = Scrambled, referring to the premise presentation order. Error bars indicate 95% confidence intervals.

Table 6

Factorial ANOVA Results for Ratings of Confidence in the Validity of Conclusions

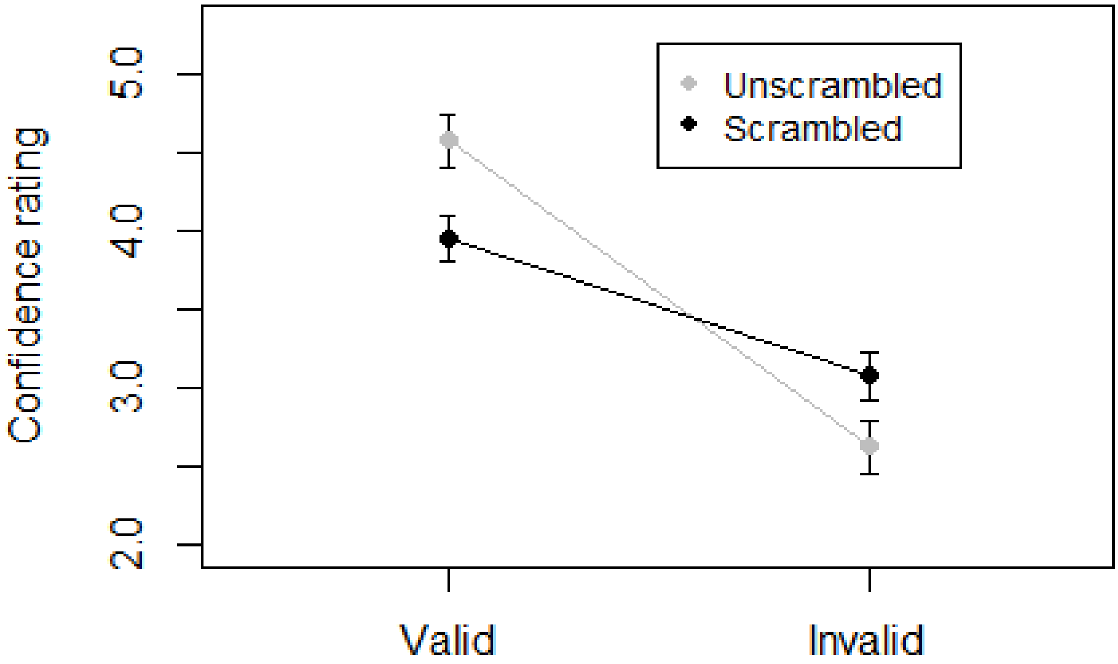
Effect	DFn	DFd	*F*	*P*	*ges*
Val	1	120	135.092	<.001*	0.272
Bel	1	120	85.615	<.001*	0.169
PremOrd	1	120	2.897	0.091	0.001
Time	1	120	0.020	0.88	0.000
Val x Bel	1	120	2.682	0.104	0.000
PremOrd x Val	1	120	56.987	<.001*	0.05
PremOrd x Bel	1	120	0.018	0.893	0.000
Time x Val	1	120	2.308	0.131	0.000
Time x Bel	1	120	0.096	0.758	0.000
PremOrd x Time	1	120	0.534	0.466	0.000
PremOrd x Val x Bel	1	120	0.137	0.712	0.000
PremOrd x Time x Val	1	120	0.007	0.934	0.000
PremOrd x Time X Bel	1	120	0.851	0.258	0.000
Time x Val x Bel	1	120	0.161	0.689	0.000
PremOrd x Time x Val x Bel	1	120	0.102	0.75	0.000

Note. Val = Validity, Bell = Believability, PremOrd = Premise order, Time = Premise presentation time, * = p <0.0, *ges* = "generalised eta squared".

It should be noted that interpretation of interaction effects identified by ANOVA (such as the validity × premise order interaction) as supporting evidence for differential effects on underlying psychological process(es) must proceed with caution (see Loftus, 1978; Wagenmakers et al., 2012). The concern is that "removable interactions" are not distinguished from "non-removable" interactions by ANOVA. Removable interactions refer to those interactions that can be undone by a monotonic transformation of data. Non-removable interactions provide more robust support for differential effects on latent processes. However, in this instance, the significant premise presentation order × validity interaction is a cross-over, non-removable, interaction (see Figure 4). This non-removable interaction suggests that there are differential effects of the premise order on participants' ability to differentiate between valid and invalid conclusions. This interaction is consistent with dual-process theories of argument evaluation, however, it may also be consistent with a single-process, independent-1D model with multiple decision threshold parameters.

Figure 4

Mean Ratings of Confidence in Validity of Conclusions



Note. Error bars show 95% confidence intervals

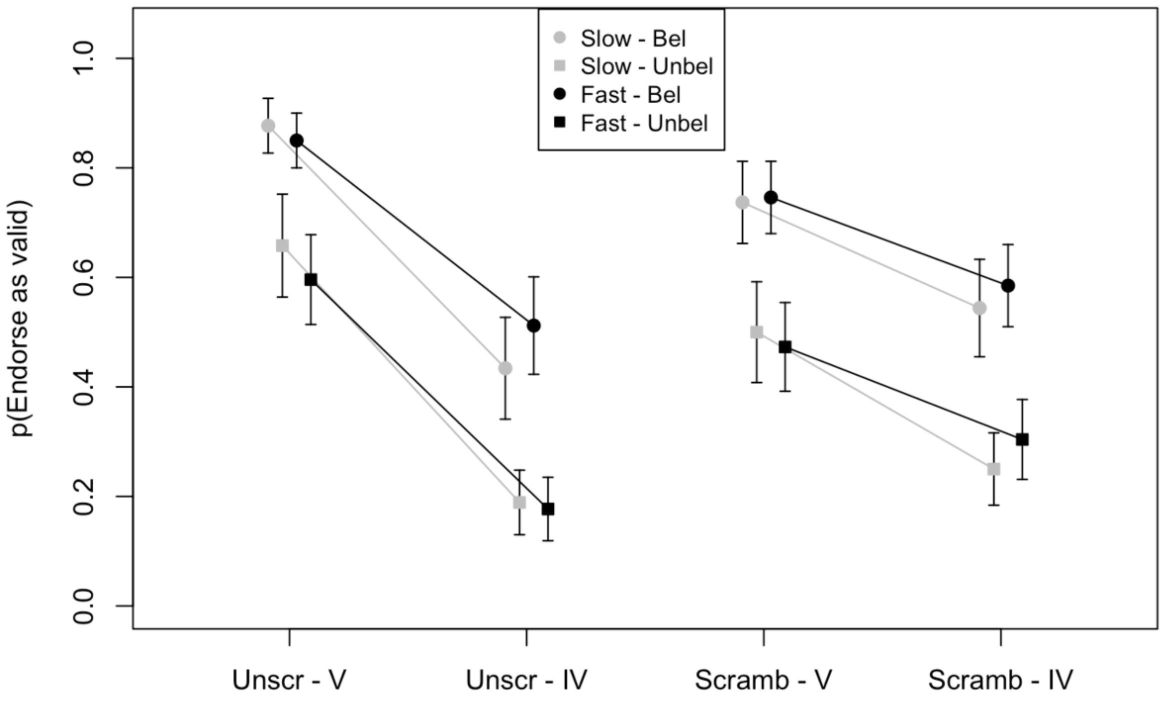
Model Testing via Signed Difference Analysis

Signed Difference Analysis was applied to test the single- and dual-process signal detection models in their most general form (i.e., with minimal assumptions about the response distributions). The aim was to test whether the ordinal pattern that is forbidden by the independent-1D model has been observed (see Figure 2a). The presence of the forbidden pattern would rule out the independent-1D model in favour of the independent-2D model. The current SDA approach and models based upon Stephens et al. (2018) require binary responses, thus, the 6-point ratings were dichotomised. Responses of 3 (“possibly valid”) or higher were considered as endorsement that the argument was valid and responses below 3 were rejections.

Figure 5 illustrates the mean endorsement rates. The mapping of the scores follows the design for SDA. The SDA “dependent variables” are displayed on the x-axis, and the “conditions” are presented as different lines. Similar effects of believability and validity, plus the interaction between validity and premise order are evident here for the binary responses, as per the original mean ratings. Most importantly, Figure 5 does not illustrate any ordinal patterns that correspond to the “double-reversed crossovers” that are forbidden by the independent-1D model. Therefore, although ANOVA identified a significant interaction between premise order and validity that would be in line with dual-process views, SDA shows that the single-process, independent-1D model cannot be rejected.

Figure 5

Mean Endorsement Rates in the Transitive Argument Evaluation Task



Note. Slow = 20 seconds premise presentation time, Fast = 10 seconds premise presentation time, V = Valid conclusions, IV = Invalid conclusions. Unscr = Unscrambled and Scramb = Scrambled, referring to the premise presentation order. Error bars indicate 95% confidence intervals.

Discussion

Summary of Findings

This study sought to rigorously test competing single-process and dual-process theories of transitive reasoning. The two main aims of this study were to test whether transitive reasoning performance is differentially affected by key experimental manipulations, as predicted by dual-process theory, and if so, test whether the effects are sufficient to rule out single-process theory. In an argument evaluation task with items that varied in validity and believability, the key manipulations were the premise integration time (10 vs 20 seconds) and working memory demands through premise order difficulty (scrambled vs. unscrambled).

The results showed that the validity of the conclusions, as well as the believability of conclusions had an effect on participants' validity ratings. Participants gave higher ratings to conclusions that were logically valid than invalid and gave higher ratings to conclusions that were believable than unbelievable. An interaction between premise presentation order and validity was observed. Participants were able to better discriminate between valid and invalid arguments when working memory demands are lower (i.e., in the unscrambled condition) than when working memory demands are high (i.e., in the scrambled condition), as predicted by dual-process accounts. However, in contrast to other predictions made by dual-process theory, the effect of time pressure on believability and validity, and the effect of premise order on believability was not observed. Signed Difference Analysis of the independent-1D and independent-2D signal detection models did not indicate compelling evidence for a dual-process explanation. The ordinal pattern of data that is forbidden by the independent-1D model was not observed, suggesting that the single-process explanation can be retained.

Discussion of Findings

The observed main effects of believability and validity provide clear evidence that people considered both logic and belief during transitive argument evaluation. This data is as expected and is consistent with the results from Andrews (2010), which the current experiment extended upon. The observed effects of both validity and believability are also consistent with previous research across various forms of deductive reasoning (see Evans, 2010). It has been long established that people tend to show belief bias, whereby deductive judgements are influenced by prior knowledge about the conclusions (Bago & De Neys, 2017). This pattern of data is explainable by dual-process theory. When evaluating arguments, it is likely that participants engaged in the associative and experience-based decision making that are key characteristics of Type 1 processing, resulting in higher ratings of conclusions that better align with prior beliefs (Evans & Stanovich, 2013). However, this pattern of data may also be explainable by single-process theory. Participants' tendency to give higher ratings to believability conclusions more than unbelievable may be the result of decision threshold differences that are determined by participants' prior knowledge of the conclusions (Stephens et al., 2019b).

The observed interaction between premise order and validity suggests that people's ability to differentiate between valid and invalid conclusions is associated with working memory resources. When evaluating conclusions in arguments that are less difficult (i.e., unscrambled condition), the cognitive demand is reduced due to premise integration strategies that can be employed, such as chaining or concatenation. These strategies assist by enabling new elements to be immediately added to a mental model upon encountering them (Andrews, 2010). In comparison, the scrambled presentation requires chunks of premises to be held active in memory as they are encountered (e.g., $d > c$ = chunk 1, $e < a$ = chunk 2), before they can be integrated as $a > b > c > d > e$ (Halford et al., 2007). This premise order and validity interaction is also consistent with the findings from Andrews (2010), and is consistent with other studies of deductive reasoning that investigated the effect of working memory through concurrent load tasks (Bago & De Neys, 2017; Hayes et al., 2018). This finding is consistent with the dual-process view that working memory is a key characteristic of Type 2 processing (Evans & Stanovich, 2013). Imposing a high working memory demand suppresses Type 2 processing, therefore, participants were not able to discriminate between valid and invalid arguments as efficiently (Evans, 2010). However, the effect of working memory demand on validity discrimination may also be consistent a single-process signal detection model that includes multiple decision criteria (Hayes et al., 2018; Stephens et al., 2018).

An important goal of this study was to more rigorously examine any differential effects of working memory demands and time pressure. The experiment was designed based on key evidence that has been used to support dual process theories, under the postulation that reasoning abilities are dependent on two distinct types of processing, and that reasoning errors like belief-bias can be made prominent when Type 2 processing is inhibited. Despise the non-removable interaction of premise order and validity, the SDA approach showed that there was no compelling evidence against the independent-1D model. Therefore, this suggests that the data are not inconsistent with the single-process view. This finding also suggests that signal detection models based on induction and deduction judgements (Stephens et al., 2018) were generalisable to a study design that focuses more directly on working memory resources, thus, paving the way for stronger tests of evidence against the 1D model.

Another important factor for dual-process theories is time pressure during argument evaluation (Evans & Stanovich, 2013). Generally, participants who operated under time pressure in experiments were more likely to endorse believable than unbelievable conclusions, regardless of their logical validity (e.g., Andrews, 2010; Evans & Curtis-

Holmes, 2005). It is typically claimed that Type 2 processing requires more time to activate and/or run its course, due to its slower, more deliberate, and effortful nature. Therefore, the belief-bias effect observed under time pressure are attributed to the inhibition of Type 2 processing (Evans & Stanovich, 2013). However, the effects of time pressure were not significant in the current study. The findings in the current study are inconsistent with Andrews' (2010) who observed the premise presentation time × validity, and premise presentation time × believability. In Andrews' (2010) study, participants were more like to accept believable conclusions than unbelievable conclusions in the shorter premise presentation condition, indicating an increased bias towards believable conclusions when less time was available. In contrast, participants in our study appeared to be engaged in a similar level of analytical processing regardless of premise presentation time. This nonsignificant effect may be problematic for the dual-process assumption that Type 1 processing tends to be faster than Type 2 processing. The nonsignificant effect of time pressure in the current study also contradicts other previous reasoning research and a methodological limitation could be the cause.

One possible methodological explanation may be that the short, 10 seconds condition did not provide enough pressure in this instance. Identical to Andrews (2010), the time limits were set at 10 and 20 seconds for the short and long conditions, respectively. However, note that in the same experiment Andrews (2010) also varied conclusion evaluation times for each condition (5 and 10 seconds), but no significant effect of conclusion evaluation time was observed. The findings of Andrews (2010) suggested that transitive reasoning performance may be more affected by time pressure at the time of premise integration rather than conclusion evaluation time, thus, the conclusion evaluation time was fixed at 5 seconds for both conditions for our experiment. Nevertheless, the implementation of the current experiment was different to that of Andrews (2010) in several other ways that could be important. Firstly, Andrews (2010) conducted the experiments using a PowerPoint presentation to display the premises, followed by participants writing down the answers on a sheet of paper. In comparison, the current study was conducted as an online experiment, with participants using a computer. This methodological difference may have resulted in a difference in participants' information processing time between the two experiments. Studies have shown that interestingly, the processing time that is required for computer-based tests is less than the processing time required for pen and paper tests (Karay et al., 2015). Therefore, less pressure may have been imposed by the "fast" condition in our computer-based experiment compared to Andrews (2010)'s fast condition via PowerPoint and paper. In addition, due to not having control over response time, we considered the possibility that participants in the short condition may have compensated by taking additional time to reflect on the premises prior to responding. However, analysis of participants' response times in both conditions supported that this was not the case. Furthermore, Karay et al. (2015) also found that the rate of guessing is higher in computer-based tests than pen and paper versions, suggesting that our data could have been more impacted by participants guessing the answers. Another point of difference between the current study and Andrews (2010) was the increased number of trials, from 16 to 32. The increased number of trials could have potentially reduced performance in validity discrimination.

Theoretical and Practical Implications

The most important goal of this experiment was to make rigorous comparison of competing single- and dual-process theories of transitive reasoning. This experiment aimed to contribute to the long theoretical debate about the cognitive processes that drive reasoning. While dual-process theories have dominated the reasoning literature, only a relatively small amount of research has directly compared single- and dual-process theories against each other (Rotello & Heit, 2009). Much of the research investigating transitive reasoning has been conducted with the aim to differentiate between variants of dual-process accounts rather than comparing single versus dual-process accounts (e.g., Andrews, 2010; Andrews & Mihelic, 2014; Banks & Hope, 2014). From a dual-process perspective, the results from the current experiment may be seen as consistent with models that suggest parallel activation of Type 1 and Type 2 processing (e.g., Evans, 2007; Sloman, 1996). These parallel processing models suggest that belief-based and analytic processing activate and operate simultaneously which could explain why the effect of time pressure was not significant; Type 2 processing was not impacted by the manipulation. At the same time, the findings in the current study could also be seen as consistent with the revised dual-process models which suggest that Type 1 processing can also generate automatic logical outputs (Bago & De Neys, 2017; Handley & Trippas, 2015). For that reason, the believability and validity of conclusions were not affected by premise presentation time as they were both able to be processed by Type 1 processing. The flexibility of these accounts means they can accommodate for a wide range of data patterns. However, this extra flexibility can prove problematic because the models are not falsifiable. Perhaps this excessive flexibility is unnecessary, and reasoning can be alternatively explained by a single underlying process of cognitive judgement, grounded in a single dimension of argument strength that is based on the consideration of a variety of cues including the believability and validity of arguments (Hayes et al., 2018).

This project has further demonstrated that single-process theories can be considered as a feasible alternative explanation of deductive judgements by formally instantiating the competing theories using signal detection models. The independent-1D and -2D models have previously been applied to a wide range of reasoning tasks and such research has consistently documented that the quantitative, single-process, independent-1D model cannot be ruled out in favour of more complex 2D alternatives (Stephens et al., 2020). The success of the 1D model thus challenges the core idea of having multiple distinct processes of argument assessment.

The approach of the current study has far-reaching implications, given that dual-process theories have been widely popularised across multiple different areas of cognitive psychology. The core assumption is that there are two types of cognitive processes, one being unconscious, uncontrollable, and unintentional; while the other that is conscious, intentional and effortful (Melnikoff & Bargh, 2018). This assumption has been used as explanatory framework for many psychological phenomena. For instance, there is reflective versus automatic response systems of emotional response coherence (Evers et al., 2014), visual versus verbal thinking in moral judgement (Amit et al., 2014), explicit versus automatic processing of social information (Evans, 2008), activation of a reflective system versus affective-automatic system in alcohol addiction (Lannoy et al., 2014), reflective and impulsive pathways in personality psychology (Perugini & Back, 2021) and declarative versus non-declarative memory and learning (McLaren et al., 2014). However, such dual-process accounts are increasingly criticised based on their: structure being a generic framework rather than a unified theory (Grayout, 2019), inconsistent distinction between two types of processing (Osman, 2004), lack of conceptual clarity and over-reliance on questionable methods and insufficient empirical evidence (Keren & Schul, 2009), and flawed logic of dissociation (Newell & Dunn, 2008). Thus, these criticisms support that single-process theories were prematurely dismissed. The current study also illustrates a general method that can be used to test competing quantitative models, which could be applied to other fields.

Signal detection theory in deductive reasoning offers a detailed account of the decision process involved by the specification of parameters that correspond to subjective argument strength or decision criteria. Although the process of how the values of these parameters are set is not explained by signal detection theory, the framework can be used to guide the development of more elaborate process models (Stephens et al., 2019b). In addition, the current model testing approach could be applied in different domains of psychology, with this approach setting an example for future studies of single-process theories and the development of models to carefully test the competing theories. If results similar to the current study are produced through the application of similar approaches in other areas where dual-process theories have been highly influential, then single-process theories can be argued as a viable alternative explanation.

The data from the current study also have several possible practical implications in everyday life. Firstly, transitive reasoning is a form of deductive reasoning, which is a skill that is highly valued in multiple domains. In research, evaluation of new scientific evidence requires systematic, logical thinking that is key to deductive reasoning; thus, deductive reasoning skills are thought to be the cornerstone of scientific research (Leighton, 2006). Deductive reasoning skills are also highly valued in clinical reasoning (Shin, 2019), for tertiary school admission (Powers & Dwyer, 2003) and are desirable in the workforce (Carnevale & Smith, 2013). Generally, deductive reasoning skills can be acquired as part of traditional content areas like mathematics and science (Leighton, 2006), and learning such skills leads to better decision making and problem solving (Klaczynski & Narasimham, 1998). Thus, training programs are implemented to strengthen reasoning skills. However, the types of training that target reasoning often apply a dual-process view via implementing the training of separate systems that target different types of thinking (e.g., Kunn et al., 2020). If dual-process views are incorrect, other training designs may be more effective.

Additionally, other areas of everyday life such as in economics (Grayout, 2020), the implementation of behavioural nudging strategies (Weijers et al., 2021), nutrition label cues (Sanjari et al., 2017), and medical diagnostics (Kuhn et al., 2020) also operate under the dual-process assumption. For instance, in behavioural nudging, Type 1 and Type 2 behavioural nudges are designed to target automatic, intuitive Type 1 reasoning or analytical, deliberate Type 2 reasoning, respectively. Behavioural nudges that target physical inactivity generally involves a point-of-decision prompts which aims to motivate people to choose the more active option (Lin et al., 2017). Type 1 nudging tends to use covert visual cues, such as painted footprints on staircases to encourage people to choose the stairs over the elevator. Type 2 nudging on the other hand, involves presenting educational information that highlights the benefits of exercise. This may be done through placing posters at the elevators or stairwells about the amount of calories they could burn and the benefits to their overall health. These kinds of real-world impact thus prove that it is imperative to understand the true underlying process(es) of reasoning.

Limitations and Future Directions

Another potential limitation of the current study is that the experiment was not conducted in a laboratory environment. In comparison to Andrews' (2010) experiment which was conducted using group administration in a classroom-like environment, our experiment utilised individual, online administration. Several potential issues can arise through experiments conducted online. Firstly, there is a limited amount of control over certain experimental conditions such as distractions, checking participants' comprehension, and monitoring adherence to instructions. Validity discrimination may thus be higher in the laboratory. Secondly, test circumventing strategies like screen capturing are not able to be properly controlled in an isolated, online experiment setting, although our results suggest that this was unlikely to be a common strategy. Lastly, the current contrasting results to Andrews (2010) may be the result of a difference in social pressure between experiments. Previous studies have shown that people perform better in a social context compared to isolated, computer-based learning (Stephens et al., 2010). Thus participants may perform more accurately in the laboratory, so there is more scope to reduce reasoning performance via manipulations such as presentation time.

Nevertheless, conducting a computer-based experiment also allowed for multiple methodological improvements such as full randomisation of content presented to participants, compared to the fixed content presentation used in Andrews' (2010) experiments. Other methodological improvements were also implemented via using nonsense terms to avoid conflicts between premise believability and the believability and validity of conclusion, and increased number of trials to collect more data. With an eye toward finding the dissociations predicted by dual-process theories, it would be beneficial for future experiments to be further conducted a) in a controlled laboratory setting, and b) with increased time pressure and increased working memory demands. The current study can guide future studies for investigating the combined effects of time and working memory on reasoning.

Conclusion

Dual-process theories have dominated the field of reasoning for many years and are increasingly criticized in the literature (e.g., Keren & Schul, 2009; Kruglasnki & Gierenzer, 2011), but dual-process models are constantly evolving in an attempt to clarify the two processing types. The success of the previously dismissed single-process theories raises the question of whether such a distinction of two processing types needs to be made at all. The current results, as well as other recent studies (e.g., Stephens et al; 2020), suggest that the effects predicted by dual-process models are also consistent with a single-process model. Overall, our data supports that the single-process accounts are still plausible contenders in this debate of reasoning processes. Determining the underlying cognitive process(es) of reasoning is theoretically and practically important, and this experiment provides groundwork for more rigorous comparison of competing single- and dual- process theories in transitive reasoning.

References

Amit, E., Gottlieb, S., & Greene, J. D. (2014). Visual versus verbal thinking and dual-process moral cognition. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 340–354). The Guilford Press.

Andrews, G. (2010). Belief-based and analytic processing in transitive inference depends on premise integration difficulty. *Memory & Cognition*, *38*(7), 928-940. https://doi.org/10.3758/mc.38.7.928

Andrews, G., & Halford, G. (1998). Children's ability to make transitive inferences: The importance of premise integration and structural complexity. *Cognitive Development*, *13*(4), 479-513. https://doi.org/10.1016/s0885-2014(98)90004-1

Andrews, G., & Halford, G. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology*, *45*(2), 153-219. https://doi.org/10.1016/s0010-0285(02)00002-6

Andrews, G., & Mihelic, M. (2014). Belief-based and analytic processing in transitive inference: Further evidence for the importance of premise integration. *Journal Of Cognitive Psychology*, *26*(5), 588-596. https://doi.org/10.1080/20445911.2014.909434

Bago, B. & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. Cognition, 158, 90-109. https://doi.org/10.1016/j.cognition.2016.10.014

Banks, A., & Hope, C. (2014). Heuristic and analytic processes in reasoning: An event-related potential study of belief bias. *Psychophysiology*, *51*(3), 290-297. https://doi.org/10.1111/psyp.12169

Bouwmeester, S., Vermunt, J., & Sijtsma, K. (2007). Development and individual differences in transitive reasoning: A fuzzy trace theory approach. *Developmental Review*, *27*(1), 41-74. https://doi.org/10.1016/j.dr.2006.08.001

Blom, S., Gillebaart, M., De Boer, F., van der Laan, N., & De Ridder, D. (2021). Under pressure: Nudging increases healthy food choice in a virtual reality supermarket, irrespective of system 1 reasoning. *Appetite*, *160*, 105116. https://doi.org/10.1016/j.appet.2021.105116

Carnevale, A., & Smith, N. (2013). Workplace basics: the skills employees need and employers want. *Human Resource Development International*, *16*(5), 491-501. https://doi.org/10.1080/13678868.2013.821267

Conway, A., Kane, M., & Engle, R. (2003). Working memory capacity and its relation to general intelligence. *Trends In Cognitive Sciences*, *7*(12), 547-552. https://doi.org/10.1016/j.tics.2003.10.005

De Neys, W. (2006). Dual processing in reasoning. *Psychological Science*, *17*(5), 428-433. https://doi.org/10.1111/j.1467-9280.2006.01723.x

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. Perspectives on Psychological Science, 7(1). https://doi.org/10.1177/1745691611429354

De Neys, W., & Van Gelder, E. (2009). Logic and belief across the lifespan: the rise and fall of belief inhibition during syllogistic reasoning. *Developmental Science*, *12*(1), 123-130. https://doi.org/10.1111/j.1467-7687.2008.00746.x

Dunn, J., & Anderson, L. (2018). Signed difference analysis: Testing for structure under monotonicity. *Journal Of Mathematical Psychology*, *85*, 36-54. https://doi.org/10.1016/j.jmp.2018.07.002

Dunn, J., & James, R. (2003). Signed difference analysis: Theory and application. *Journal Of Mathematical Psychology*, *47*(4), 389-416. https://doi.org/10.1016/s0022-2496(03)00049-x

Dunn, J., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, *95*(1), 91-101. https://doi.org/10.1037/0033-295x.95.1.91

Evans, J. (2003). In two minds: dual-process accounts of reasoning. *Trends In Cognitive Sciences*, *7*(10), 454-459. https://doi.org/10.1016/j.tics.2003.08.012

Evans, J. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, *13*(4), 321-339. https://doi.org/10.1080/13546780601008825

Evans, J. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. Annual Review of Psychology, 59, 255-278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. (2010). Intuition and reasoning: a dual-process perspective. *Psychological Inquiry*, *21*(4), 313-326. https://doi.org/10.1080/1047840x.2010.521057

Evans, J., Barston, J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295-306. https://doi.org/10.3758/bf03196976

Evans, J., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*(4), 382-389. https://doi.org/10.1080/13546780542000005

Evans, J., Handley, S., Neilens, H., & Over, D. (2010). The influence of cognitive ability and instructional set on causal conditional inference. *Quarterly Journal Of*

Experimental Psychology, *63*(5), 892-909. https://doi.org/10.1080/17470210903111821

Evans, J., & Stanovich, K. (2013). Dual-process theories of higher cognition. *Perspectives On Psychological Science*, *8*(3), 223-241. https://doi.org/10.1177/1745691612460685

Evers, C., Hopp, H., Gross, J., Fischer, A., Manstead, A., & Mauss, I. (2014). Emotion response coherence: A dual-process perspective. *Biological Psychology*, *98*, 43-49. https://doi.org/10.1016/j.biopsycho.2013.11.003

Goel, V., Navarrete, G., Noveck, I., & Prado, J. (2017). Editorial: the reasoning brain: the interplay between cognitive neuroscience and theories of reasoning. *Frontiers In Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00673

Grayot, J. (2019). Dual process theories in behavioural economics and neuroeconomics: a critical review. *Review Of Philosophy And Psychology*, *11*(1), 105-136. https://doi.org/10.1007/s13164-019-00446-9

Halford, G., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: a new hypothesis. *Trends In Cognitive Sciences*, *11*(6), 236-242. https://doi.org/10.1016/j.tics.2007.04.001

Halford, G., Wilson, W., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioural And Brain Sciences*, *21*(6), 803-831. https://doi.org/10.1017/s0140525x98001769

Handley, S. J., & Trippas, D. (2015). Dual processes and the interplay between knowledge and structure: A new parallel processing model. Psychology of Learning and Motivation, 62, 33-58. https://doi.org/10.1016/bs.plm.2014.09.002

Hayes, B., Dunn, J., Joubert, A., & Taylor, R. (2016). Comparing single- and dual-process models of memory development. *Developmental Science*, *20*(6). https://doi.org/10.1111/desc.12469

Hayes, B., Stephens, R., Ngo, J., & Dunn, J. (2018). The dimensionality of reasoning: Inductive and deductive inference can be explained by a single process. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *44*(9), 1333-1351. https://doi.org/10.1037/xlm0000527

Howarth, S., Handley, S., & Walsh, C. (2016). The logic-bias effect: The role of effortful processing in the resolution of belief–logic conflict. *Memory & Cognition*, *44*(2), 330-349. https://doi.org/10.3758/s13421-015-0555-x

Hunt, E., & Madhyastha, T. (2012). Cognitive demands of the workplace. *Journal Of Neuroscience, Psychology, And Economics*, *5*(1), 18-37. https://doi.org/10.1037/a0026177

Johnson-Laird, P. (1999). Deductive Reasoning. *Annual Review Of Psychology*, *50*(1), 109-135. https://doi.org/10.1146/annurev.psych.50.1.109

Karay, Y., Schauber, S., Stosch, C., & Schüttpelz-Brauns, K. (2015). Computer Versus Paper—Does It Make Any Difference in Test Performance?. *Teaching And Learning In Medicine*, *27*(1), 57-62. https://doi.org/10.1080/10401334.2014.979175

Keren, G., & Schul, Y. (2009). Two Is Not Always Better Than One. *Perspectives On Psychological Science*, *4*(6), 533-550. https://doi.org/10.1111/j.1745-6924.2009.01164.x

Klaczynski, P., & Narasimham, G. (1998). Development of scientific reasoning biases: Cognitive versus ego-protective explanations. *Developmental Psychology*, *34*(1), 175-187. https://doi.org/10.1037/0012-1649.34.1.175

Klauer, K., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, *107*(4), 852-884. https://doi.org/10.1037/0033-295x.107.4.852

Kuhn, J., van den Berg, P., Mamede, S., Zwaan, L., Diemers, A., Bindels, P., & van Gog, T. (2020). Can we teach reflective reasoning in general-practice training through example-based learning and learning by doing?. *Health Professions Education*, *6*(4), 506-515. https://doi.org/10.1016/j.hpe.2020.07.004

Kruglanski, A.W. & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. Psychological Review, 118(1), 97-109. https://doi.org/10.1037/a0020762

Lannoy, S., Billieux, J., & Maurage, P. (2014). Beyond Inhibition: A Dual-Process Perspective to Renew the Exploration of Binge Drinking. *Frontiers In Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00405

Lazareva, O. (2012). Transitive inference in nonhuman animals. *Oxford Handbooks Online*. https://doi.org/10.1093/oxfordhb/9780195392661.013.0036

Leighton, J. (2006). Teaching and Assessing Deductive Reasoning Skills. *The Journal Of Experimental Education*, *74*(2), 107-136. https://doi.org/10.3200/jexe.74.2.107-136

Lin, Y., Osman, M., & Ashcroft, R. (2017). Nudge: Concept, Effectiveness, and Ethics. *Basic And Applied Social Psychology*, *39*(6), 293-306. https://doi.org/10.1080/01973533.2017.1356304

Loftus, G. R. (1978). On interpretation of interactions. Memory & Cognition, 6, 312-319. https://doi.org/10.3758/BF0319746

Markovits, H., & Barrouillet, P. (2004). : Introduction: Why is understanding the development of reasoning important?. *Thinking & Reasoning*, *10*(2), 113-121. https://doi.org/10.1080/13546780442000006

McLaren, I., Forrest, C., McLaren, R., Jones, F., Aitken, M., & Mackintosh, N. (2014). Associations and propositions: The case for a dual-process account of learning in humans. *Neurobiology Of Learning And Memory*, *108*, 185-195. https://doi.org/10.1016/j.nlm.2013.09.014

Melnikoff, D., & Bargh, J. (2018). The Mythical Number Two. *Trends In Cognitive Sciences*, *22*(4), 280-293. https://doi.org/10.1016/j.tics.2018.02.001

Miulescu, T., Chraif, M., Aniţei, M., Iliescu, A., & Bârcă, A. (2012). The influence of faculty specialization on deductive reasoning in young Romanian students. *Procedia - Social And Behavioral Sciences*, *33*, 1063-1066. https://doi.org/10.1016/j.sbspro.2012.01.285

Morsanyi, K., McCormack, T., & O'Mahony, E. (2017). The link between deductive reasoning and mathematics. *Thinking & Reasoning*, *24*(2), 234-257. https://doi.org/10.1080/13546783.2017.1384760

Newell, B., & Dunn, J. (2008). Dimensions in data: testing psychological models using state-trace analysis. *Trends In Cognitive Sciences*, *12*(8), 285-290. https://doi.org/10.1016/j.tics.2008.04.009

Niu, W., Zhang, J., & Yang, Y. (2007). Deductive reasoning and creativity: a cross-cultural study. *Psychological Reports*, *100*(2), 509-519. https://doi.org/10.2466/pr0.100.2.509-519

Oakhill, J., & Johnson-Laird, P. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *The Quarterly Journal Of Experimental Psychology Section A*, *37*(4), 553-569. https://doi.org/10.1080/14640748508400919

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends In Cognitive Sciences*, *5*(8), 349-357. https://doi.org/10.1016/s1364-6613(00)01699-5

Osman, M. (2004). An evaluation of dual-process theories of reasoning. Psychonomic Bulletin & Review, 11(6), 988-1010. https://doi.org/10.3758/bf03196730

Pastore, R., & Scheirer, C. (1974). Signal detection theory: Considerations for general application. *Psychological Bulletin*, *81*(12), 945-958. https://doi.org/10.1037/h0037357

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. Cognitive Psychology, 80, 34-72. https://doi.org/10.1016/j.cogpsuch.2015.05.001

Perugini, M., Hagemeyer, B., Wrzus, C., & Back, M. (2021). Dual process models of personality. *The Handbook Of Personality Dynamics And Processes*, 551-577. https://doi.org/10.1016/b978-0-12-813995-0.00020-0

Phillips, S., Wilson, W., & Halford, G. (2009). What do transitive inference and class inclusion have in common? categorical (co)products and cognitive development. *Plos Computational Biology*, *5*(12), e1000599. https://doi.org/10.1371/journal.pcbi.1000599

Powers, D. E., & Dwyer, C. A. (2003). *Toward specifying a construct of reasoning. Princeton, NJ:* Educational Testing Services.

Prado, J., Chadha, A., & Booth, J. (2011). The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. *Journal Of Cognitive Neuroscience*, *23*(11), 3483-3497. https://doi.org/10.1162/jocn_a_00063

Rips, L. (2001). Two kinds of reasoning. *Psychological Science*, *12*(2), 129-134. https://doi.org/10.1111/1467-9280.00322

Rotello, C., & Heit, E. (2009). Modeling the effects of argument length and validity on inductive and deductive reasoning. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *35*(5), 1317-1330. https://doi.org/10.1037/a0016648

Sanjari, S. S., Jahn, S., & Boztug, Y. (2017). Dual-process theory and consumer response to front-of-package nutrition label formats. *Nutrition Reviews*, 75(11), 871-882. https://doi.org/10.1093/nutrit/nux043

Shin, H. (2019). Reasoning processes in clinical reasoning: from the perspective of cognitive psychology. *Korean Journal Of Medical Education*, *31*(4), 299-308. https://doi.org/10.3946/kjme.2019.140

Sloman, S.A. (1996). The empirical case for two systems of reasoning. Psychological Bulletin, 119(1), 3-22. https://doi.org/10.1037/0033-2909.119.1.3

Stanovich, K., & West, R. (2008). On the relative independence of thinking biases and cognitive ability. *Journal Of Personality And Social Psychology*, *94*(4), 672-695. https://doi.org/10.1037/0022-3514.94.4.672

Stephens, R., Dunn, J., & Hayes, B. (2018). Are there two processes in reasoning? The dimensionality of inductive and deductive inferences. *Psychological Review*, *125*(2), 218-244. https://doi.org/10.1037/rev0000088

Stephens, R., Dunn, J., & Hayes, B. (2019). Belief bias is response bias: Evidence from a two-step signal detection model. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *45*(2), 320-332. https://doi.org/10.1037/xlm0000587

Stephens, R.G., Dunn, J.C., Hayes, B.K., & Kalish, M.L. (2020). A test of two processes: The effect of training on deductive and inductive reasoning. Cognition, 199. https://doi.org/10.1016/j.cognition.2020.104223

Stephens, R., Matzke, D., & Hayes, B. (2019). Disappearing dissociations in experimental psychology: Using state-trace analysis to test for multiple processes. *Journal Of Mathematical Psychology*, *90*, 3-22. https://doi.org/10.1016/j.jmp.2018.11.003

Stephens, R., Prefors, A., & Navarro, D. (2010). Social Context Effects on the Impact of Category Labels. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 32. Retrieved from https://escholarship.org/uc/item/1063k6j5

Stupple, E., & Ball, L. (2008). Belief–logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning*, *14*(2), 168-181. https://doi.org/10.1080/13546780701739782

Todd, J., Andrews, G., & Conlon, E. (2019). Relational thinking in later adulthood. *Psychology And Aging*, *34*(4), 486-501. https://doi.org/10.1037/pag0000346

Wagenmakers, E. J., Krypotos, A. M., Criss, A. H., & Iverson, G. (2012). On the interpretation of removable interactions: A survey of the field 33 years after Loftus. Memory & Cognition, 40(2), 145-160. https://doi.org/10.3758/s13421-011-0158-0

Weijers, R., de Koning, B., & Paas, F. (2020). Nudging in education: from theory towards guidelines for successful implementation. *European Journal Of Psychology Of Education*, *36*(3), 883-902. https://doi.org/10.1007/s10212-020-00495-0

Wixted, J., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*(2), 262-276. https://doi.org/10.1037/a0035940

Appendix A

Table A
Means of Unbelievability Ratings of Conclusion in the Pilot Study

Item	Believable Form	*M*	*SD*	Unbelievable Form	*M*	*SD*
1	Dogs are bigger than mice	1.24	0.78	Mice are bigger than dogs	4.72	0.65
2	Rabbits are faster than turtle	1.45	0.79	Turtles are faster than rabbits	4.04	1.05
3	Camels are taller than donkeys	1.61	0.80	Donkeys are taller than camels	4.19	0.86
4	Sharks are bigger than tuna	1.67	0.67	Tuna are bigger than sharks	3.9	0.99
5	Monkeys are hairier than salmon	1.16	0.58	Salmon are hairier than monkeys	4.85	0.36
6	Wood is harder than cushions	1.18	0.72	Cushions are harder than wood	4.28	1.20
7	Sofas are softer than floors	1.22	0.42	Floors are softer than sofas	4.08	0.78
8	Netballs are bigger than tennis balls	1.21	0.74	Tennis balls are bigger than netballs	4.88	0.33
9	Sweaters are warmer than t-shirts	1.13	0.33	T-shirts are warmer than sweaters	4.65	0.56
10	Textbooks are longer than magazines	1.6	0.7	Magazines are longer than textbooks	4.24	0.79
11	Cakes are sweeter than lemons	1.38	0.81	Lemons are sweeter than cakes	4.40	0.89
12	Subway is healthier than McDonalds	2.03	0.72	McDonalds is healthier than Subway	4.06	0.8
13	Apples are crunchier than bananas	1.62	1.13	Bananas are crunchier than apples	4.77	0.52
14	Coconuts are bigger than plums	1.37	0.49	Plums are bigger than coconuts	4.41	0.76
15	Marshmallows are softer than crisps	1.27	0.82	Crisps are softer than marshmallows	4.78	0.48
16	Yelling is louder than talking	1.35	0.88	Talking is louder than yelling	4.25	0.94
17	Days are longer than minutes	1.86	0.48	Minutes are longer than days	4.74	0.56
18	Jet planes are louder than vacuums	1.12	0.34	Vacuums are louder than jet planes	4.7	0.52
19	Hills are higher than plains	1.8	0.96	Plains are higher than hills	4.14	0.93
20	Meters are longer than centimetres	1.65	1.28	Centimetres are longer than meters	4.67	0.70
21	Town houses are taller than cottage houses	1.94	0.75	Cottage houses are taller than townhouses	3.88	0.90

22	Decades are longer than months	1.29	0.8	Months are longer than decades	4.74	0.56
23	Morse Code is older than email	1.33	0.88	Email is older than Morse Code	4.36	1.11
24	Tesla is more expensive than Mazda	1.48	0.73	Mazda is more expensive than Tesla	4.22	0.97
25	Planets are bigger than molecules	1.42	1.06	Molecules are bigger than planets	4.5	1.02
26	Hammers are heavier than pencils	1.14	0.58	Pencils are heavier than hammers	4.62	0.82
27	Horses are stronger than cats	1.35	0.84	Cats are stronger than horses	4.42	0.59
28	Snow is colder than fire	1.13	0.34	Fire is colder than snow	4.85	0.36
29	Knives are sharper than forks	1.54	0.77	Forks are shaper than knives	3.93	0.92
30	Apes are hairier than lizards	1.19	0.75	Lizards are hairier than apes	4.66	0.62
31	Water bottles are lighter than bricks	1.92	1.06	Bricks are lighter than water bottles	4.24	0.89
32	Instagram in newer than MySpace	1.6	1.15	MySpace is newer than Instagram	4.39	0.75

Appendix B

The current thesis focuses on the scramble and unscrambled conditions as the dependent variables of SDA, while future publication of the work will present both analyses.

ASPREDICTED

Wharton CREDIBILITY LAB
UNIVERSITY *of* PENNSYLVANIA

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Transitive reasoning – Experiment 1 – Scott & Stephens - 2021 (#67548)

Created: 06/02/2021 07:39 PM (PT)

This is an anonymized copy (without author names) of the pre-registration. It was created by the author(s) to use during peer-review.
A non-anonymized version (containing author names) should be made available by the authors when the work it supports is made public.

---

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Do believability, validity, and problem complexity (premise order) in transitive reasoning have differential effects on validity ratings when the premise presentation duration is short vs. long (10 vs 20 secs)?
On the other hand, do believability, validity, and presentation duration have differential effects on ratings for more versus less complex problems?
If so, is a "single process" account of reasoning inconsistent with these effects?

3) Describe the key dependent variable(s) specifying how they will be measured.
Endorsement rates of conclusions/confidence ratings for the different types of arguments

4) How many and which conditions will participants be assigned to?
A mixed design:
Within participants - 2 (validity of conclusion) x 2 (believability of conclusion) x 2 (premise order)
Between participants - 2 (premise presentation time)

- Valid vs invalid conclusion
- Believable vs not believable conclusion
- Short presentation time (10 seconds) vs long presentation time (20 seconds)
- Ordered premise presentation vs scrambled premise presentation

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
- Mixed ANOVA / LMEs
- Signed difference analysis to test the "independent-1D" signal detection model. Dependent variables for this analysis will be validity*presentation-time and validity* premise-order

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
If there are observable patterns of inattention/boredom (ie. Answering with the same response for the majority of trails (more than 70%)

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
100 (50 vs 50)

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
No

Version of AsPredicted Questions: 2.00

Available at https://aspredicted.org/TCD_AZU

Appendix C

Table C

List of the 4 Fixed Scrambling Structures

Order	Structure	Example	Elements
1	*a* > *b*	Elephants are bigger than dogs	*a* = elephants
	d < *c*	Mice are smaller than zoots	*b* = dogs
	e < *d*	Worms are smaller than mice	*c* = zoots
	b > *c*	Dogs are bigger than zoots	*d* = mice
			e = worms
2	*d* > *e*	Magazines are longer than leaflets	*a* = encyclopedias
	b > *c*	Textbooks are longer than fronnets	*b* = textbooks
	d < *c*	Magazines are shorter than fronnets	*c* = fronnets
	b < *a*	Textbooks are shorter than encyclopedias	*d* = magazines
			e = leaflets
3	*c* > *d*	Glugies are sweeter than lemons	*a* = chocolates
	a > *b*	Chocolates are sweeter than cakes	*b* = cakes
	c < *b*	Glugies are less sweet than cakes	*c* = glugies
	e < *d*	Radishes are less sweet than lemons	*d* = lemons
			e = radishes
4	c < b	Zids are shorter than meters	*a* = kilometres
	d > e	Centimetres are longer than millimetres	*b* = meters
	b < a	Meters are shorter than kilometres	*c* = zids
	c > d	Zids are longer than centimetres	*d* = centimetres
			e = millimetres