Single or dual processing in reasoning development? An application of state-trace analysis to a systematic database of studies

Minling Tan

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

School of Psychology

University of Adelaide

September 2020

Word Count: 9392

Table of Contents

List of Figures

List of Tables

Abstract

Influential dual-process theories of higher cognition posit that two qualitatively different processes underlie human reasoning. In contrast, single-process theories postulate that reasoners draw upon common cognitive mechanisms when making inferences. To test the competing theories, a rigorous method – state-trace analysis – has been proposed and proven to be a useful tool for beginning to diagnose the number of underlying psychological processes. This approach has been previously applied in exclusively adult based populations, suggesting that single-process theories of reasoning cannot be ruled out. However, to date it remains unclear whether such results hold across the period of child development. Therefore, the current study aimed to build a database of published developmental reasoning studies and to re-evaluate the data using state-trace analysis, to determine whether they best support the single-process or dual-process accounts. An electronic search of the PsycINFO and Scopus databases was undertaken to obtain empirical studies that have applied dual-process theories to examine reasoning in children or young adolescents (6-15 years). Two screening processes identified 10 papers that provided suitable summary data, forming a database of 78 datasets. State-trace analysis was applied to each dataset. Much of the developmental reasoning data were found to be consistent with a single-process account with one underlying latent variable, thus providing limited evidence for dual-process accounts of reasoning. More targeted experimental design and more stringent statistical tools are recommended for future research, to better understand the cognitive mechanisms underlying reasoning and its development.

Keywords: reasoning, dual-process theories, child development, database, state-trace analysis

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 2020

Contribution Statement

The research question and knowledge gap were identified in consultation with my supervisor, Dr Rachel Stephens. Further, training in the methodology of database analysis as well as the statistical approach of state-trace analysis used in this thesis was provided to me by Dr Stephens, including the provision of the state-trace analysis code for use in MATLAB. The literature search strategy was devised in collaboration with both Dr Rachel Stephens and Dr Mark Kohler. In building the developmental reasoning database, I conducted the electronic database search and undertook the screening processes. Dr Stephens helped check a subset of 20 papers to assist in finalising the inclusion decisions. I constructed the 78 individual datasets for state-trace analysis, with guidance from Dr Stephens. I conducted all the data analyses, generated all the figures, and wrote up all aspects of the thesis.

Acknowledgements

I would like to acknowledge the exceptional support I have received from my two supervisors, Dr Rachel Stephens and Dr Mark Kohler, throughout the course of the project. Thank you both for taking this project under your wings, sharing your precious time and expertise, and offering timely and constructive feedback throughout. You have helped increase my confidence and skills in psychological research and academic writing. Year 2020 has been a difficult year for me. I am forever indebted to you because without your tireless support and encouragement, I might have discontinued my study and will not be completing this degree. Special thank you to Dr Rachel Stephens. You have been a marvellous supervisor guiding my journey in the study of reasoning. Your attention to detail, patience, and warmth have made my learning experience enjoyable. I cannot thank you enough for all your help. Thank you again to Dr Mark Kohler. I have been fortunate to benefit from your passion and knowledge in child development, and I have enjoyed our virtual and face-to-face meetings.

Chapter 1: Introduction

1.1 Overview

Reasoning is a fundamental cognitive ability that develops across childhood and distinguishes humans from other species (Markovits, 2017). In its most general sense, reasoning can be defined as a mental activity that involves drawing an inference from evidence or in accordance with a set of rules that govern the type of inference (Overton, 1990). The associated questions of how people reason and how reasoning develops have been very important research themes in the field of psychology.

In exploring these questions, a key observation is that adults are prone to making systematic reasoning errors (e.g., Evans, Barston, & Pollard, 1983; Kahneman, 2011). Consider, for example, an *argument evaluation task*, in which participants are asked to judge whether an argument's conclusion is deductively valid (Markovits, de Chantal, Brisson, & Gagnon-St-Pierre, 2019), such as:

If a ball is thrown into a window, then the window will break.

The window is broken

Therefore, a ball was thrown into the window.

In this example, the conclusion is deductively invalid because the window can be broken by means other than a ball, but the conclusion is believable based on background knowledge about the argument content. Therefore, if people correctly evaluate the conclusion based on whether it follows logically from the premises, they will decide that it is invalid. However, people often instead show *belief bias* – they judge the conclusion to be valid if its believability confirms with their prior belief (see Evans, 2008) – especially when under time pressure or when available cognitive resources are limited (also see Evans & Stanovich, 2013). Such evidence suggests that there are two main processes of reasoning, with one being reflective and the other intuitive (Evans & Stanovich, 2013). Nevertheless, it remains unclear what cognitive mechanisms underlie these reasoning responses – whether they reflect the operation of a common core process or two qualitatively different types of processes (Stephens, Dunn, Hayes, & Kalish, 2020).

While developmental psychologists have predicted and often observed improvement in the ability to reason from childhood to adulthood (e.g., Inhelder & Piaget, 1958; Venet & Markovits, 2001), the literature has presented a complex developmental picture. Studies have provided evidence for unexpected developmental reversals, showing that children can sometimes reason more logically than do adults (e.g., Morsanyi & Handley, 2008), as well as evidence for U-shaped developmental functions, revealing that reasoning responses become less normative with age before turning more normative again (e.g., Chiesi, Gronchi, & Primi, 2008). Empirical data from developmental reasoning studies can therefore be valuable when testing competing theories that were originally based on adult reasoning, to extend their validity, or to further inform them. However, seldom have developmental data been used to test or refine those theories – including testing for multiple reasoning processes across the childhood period (Barrouillet, 2011).

1.2 Dual-process theories of reasoning

A dominant framework for understanding different ways of making inferences is the *dual-process* model. Many dual-process theories of reasoning have been proposed (e.g., Evans, 2008; Kahneman & Frederick, 2002; Sloman, 1996; Stanovich, 2004, 2011). These theories posit that two qualitatively distinct types of cognitive processes underlie reasoning responses, having in common a distinction between processes that are fast and intuitive and those that are slow and deliberative (see Evans & Stanovich, 2013).

The two types of processes are often referred to as Type 1 and Type 2 processing[1] respectively. The central feature of Type 1 processes is their autonomy – their execution makes minimal demands on working memory (see Evans & Stanovich, 2013). Type 1 processing thus tends to be rapid, has high processing capacity, and yields responses biased by prior knowledge. In contrast, Type 2 processing is not autonomous but seen as reflective. It involves effortful hypothetical thinking. Its defining feature is cognitive decoupling – the ability to sustain the decoupling of secondary representations. Type 2 processing thus tends to be slow, loads heavily on working memory, is related to individual differences in cognitive ability, and leads to normative responses (Evans & Stanovich, 2013).

Despite the common recognition of two distinct processes that compete for control of the reasoning response, there are differences in opinion regarding how these processes interact. For instance, some theorists prefer a default-interventionist dual-process account, assuming that Type 1 processing will quickly produce a default response unless intervened by Type 2 processing (e.g., Evans, 2011; Evans & Stanovich, 2013). In contrast, others favour a parallel-competitive dual-process account, suggesting that Type 1 and Type 2 processing occur simultaneously at the problem onset and operate in parallel (e.g., De Neys, 2012; Handley & Trippas, 2015).

1.3 Single-process theories of reasoning

Although dual-process models currently dominate the reasoning literature, they have not gone unchallenged (e.g., Osman, 2013; Stephens, Dunn, & Hayes, 2018). Rather than having two different reasoning processes, *single-process* theories suppose that reasoners draw upon a common pool of fundamental cognitive processes when making inferences (cf. Keren

[1] Dual-process theories (e.g., Evans, 2011; Stanovich, 2011) have replaced the System 1/2 terminology with the Type 1/2 processing distinction. The terms System 1 and System 2 are problematic because they may incorrectly suggest singularity and the operation of two distinct brain systems while the two systems are assumed to include a variety of processes (Stanovich & Toplak, 2012).

& Schul, 2009; Kruglanski & Gigerenzer, 2011; Lassiter & Goodman, 2015; Rips, 2001). Importantly, single-process theorists argue against the qualitative dichotomy between Type 1 and Type 2 processing. Keren and Schul (2009) for example, noted that dimensions (e.g., judgement speed, resource-dependence) assumed to distinguish the two processes are not dichotomous but continuous (also see Kruglanski, 2013). Kruglanski and Gigerenzer (2011) have furthermore argued that the same rules can underlie both intuitive and deliberate judgements, proposing that the set of applicable rules are constrained by the task itself and a reasoner's memory, while the final selection of a rule is determined by the perceived ecological rationality of the rule and the reasoner's processing potential. More recently, based on a signal detection framework, a body of empirical work provides evidence for a single-process reasoning model of the argument evaluation task, which assumes a single latent dimension for evaluating argument strength and allows response criteria (e.g., for endorsing a conclusion as valid) to shift across variants of the task (e.g., Hayes, Wei, Dunn, & Stephens, 2020; Stephens et al., 2020).

1.4 Dual-process theories – evidence and reliance on dissociation logic

An important pattern of data for inferring the number of underlying cognitive processes to account for a psychological phenomenon has been dissociation – an observation that the performance on a task is differentially affected by one or more factors (Newell & Dunn, 2008). Accordingly, dual-process theories have advanced on the basis of dissociations in data and in particular on interaction effects revealed by statistical models such as analysis of variance (ANOVA) (Newell & Dunn, 2008). Key evidence for dual processes in adult reasoning has accumulated from experimental tasks that contrast a heuristic response option thought to be produced by intuitive Type 1 processing with a normative response option thought to involve reflective Type 2 processing, such as base-rate neglect problems, conjunction fallacy problems, or the belief bias (or argument evaluation) paradigm (e.g., De

Neys, 2006; De Neys & Glumicic, 2008). For instance, in the belief bias paradigm, participants make decisions about the validity of syllogisms that vary in validity (valid or invalid; this defines normative response options) and believability (believable or unbelievable; this defines heuristic response options). Typically, studies adopting the paradigm find a belief by logic interaction, whereby belief bias is much more marked on invalid syllogisms than valid syllogisms (e.g., Evans et al., 1983). Such an effect has been characterised as involving logic-based processes and belief-based processes, which supports dual-process accounts of reasoning (Evans et al., 1983).

Stronger support for dual processes in reasoning has been provided by belief-bias studies with experimental manipulations designed to affect one type of processing while leaving the other unaffected. Remember that dual-process theories postulate that Type 1 processes tend to be fast and autonomous, whereas Type 2 processes tend to be slow and working memory-intensive. Thus, limited time or increased working memory load should cue Type 1 processing and inhibit Type 2 processing. As predicted by dual-process theories, studies have consistently found a larger effect of belief and a smaller effect of logic on reasoning responses in participants under time pressure (e.g., Evans & Curtis-Holmes, 2005) and under concurrent working memory loads (e.g., De Neys, 2006). Conversely, a smaller effect of belief and a larger effect of logic on reasoning responses were found in tasks with strict instructions to reason deductively, which are thought to selectively increase Type 2 processing effort (e.g., Evans, Handley, Neilens, & Over, 2010). Crucially, these interaction effects obtained empirically have been interpreted as strong evidence for dual-process accounts of reasoning (see Evans & Stanovich, 2013 for a review).

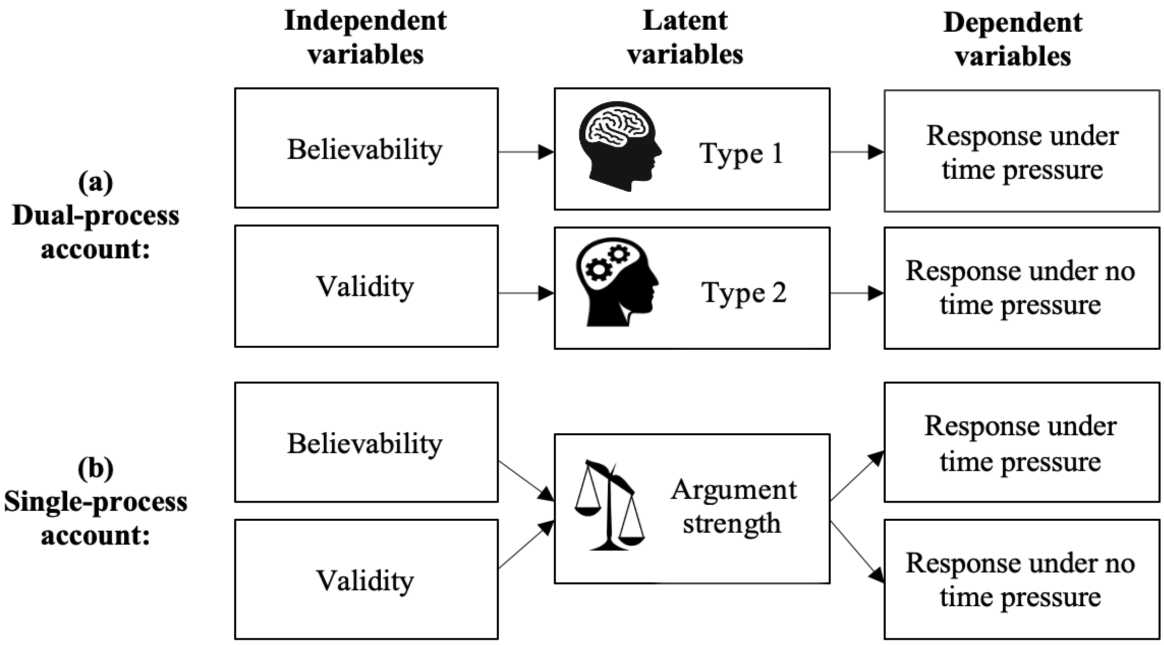

Figure 1. Two different models of two independent variables (e.g., believability and validity in the argument evaluation task) on the two dependent variables (e.g., proportion correct on validity assessments under time pressure and no time pressure). (a) A possible dual-process model that assumes the performance on the two dependent variables is driven by two underlying latent variables, reflecting the output of Type 1 versus Type 2 processing. (b) A single-process model that assumes the performance on both dependent variables is driven by a single underlying latent variable, such as the subjective strength of an argument.

However, reliance on the standard analysis of interactions to make inferences about multiple underlying cognitive processes can be problematic (Loftus, 1978; Wagenmakers, Krypotos, Criss, & Iverson, 2012). To help illustrate the problem, a diagram of the competing theories is shown in Figure 1, using the argument evaluation or belief bias task as an example. A dual-process model is in theory a multiple-parameter model in which the effects of independent variables (e.g., argument validity and believability) on two manifest dependent variables (e.g., observed validity assessments under time pressure versus no time pressure) are mediated by two latent psychological variables (Type 1 versus Type 2 processing) (Dunn, Kalish, & Newell, 2014). This can be contrasted with the simplest single-process model with a single latent variable (e.g., subjective strength of an argument).

A critical consideration is that the precise functional mapping between the dependent variables (which are measured) and the latent psychological variable(s) (which are inferred) is unknown (Loftus, 1978). Most importantly, the rejection of a single-process model on the basis of dissociations (i.e., interaction effects identified by linear statistical models such as ANOVA) usually depends on the strong and unwarranted assumption that the latent variable has the same *linear* mapping onto the observed responses for each dependent variable (see Stephens, Matzke, & Hayes, 2019 for a more detailed discussion). However, as demonstrated by Loftus (1978) and Wagenmakers et al. (2012), if these mappings are instead *non-linear*, interaction effects may actually reflect equal shifts in a latent variable across conditions (e.g., across the argument validity/believability conditions). Thus, dissociations can in fact be consistent with a single-process model.

1.5 State-trace analysis and disappearing dissociations

To avoid over-interpreting evidence as supporting multiple-process models, state-trace analysis (STA) has been suggested to replace the inferential role of dissociations (Newell & Dunn, 2008). State-trace analysis (Bamber 1979; Dunn, 2008) is a general method to help determine the number of latent variables that mediate the effects of one or more independent variables on a set of dependent variables. This approach overcomes the flaws of dissociation logic in that it subsumes other patterns of data (e.g., single dissociation or crossed dissociation – see Dunn & Kirsner, 1988) as special cases. Also, it makes a milder and more realistic assumption that each dependent variable is an unknown (and potentially different) *monotonic* function of the latent variable(s) and not necessarily a linear function. The assumption of monotonicity implies that changes in the latent variable are never followed by changes in the dependent variable in the opposite direction; that is, as the latent variable increases or decreases, the dependent variable always increases or decreases in the same direction or stays the same (Dunn et al., 2014). Notably, state-trace analysis has been increasingly used in reasoning studies and other areas of cognitive science and proven to be a useful tool in diagnosing the underlying dimensionality of experimental tasks that test rival single- and multiple-process models (Dunn et al., 2014).

An important tool in STA is the state-trace plot (Dunn, 2008). By producing a plot of one dependent variable (e.g., conclusion acceptance rates under standard deductive or "categorical" constructions in the argument evaluation task) as a function of another dependent variable (e.g., ratings of how likely a conclusion is under alternative "probabilistic" instructions), one can infer the number of latent variables (or psychological processes) required to account for the observed data (see Figures 2a and 2b for examples). A monotonic plot indicates that the data are consistent with the operation of one underlying latent variable (e.g., Figure 2a), suggesting that there is no need to posit multiple processes. In contrast, a plot that shows non-monotonic discontinuities suggests the operation of more than one latent variable (e.g., Figure 2b), which may support a dual-process account. The general rationale is that if there is only one latent variable (e.g., subjective argument strength), a positive change in this variable that produces an *increase* in one of the dependent variables cannot simultaneously produce a *decrease* in the other dependent variable. In addition to the examination of state-trace plots, Dunn and Kalish (2018) have developed a statistical test that identifies the best-fitting monotonic state-trace curve (see examples in Figure 2), which can test whether there are significant departures from the curve.

(a) (b)

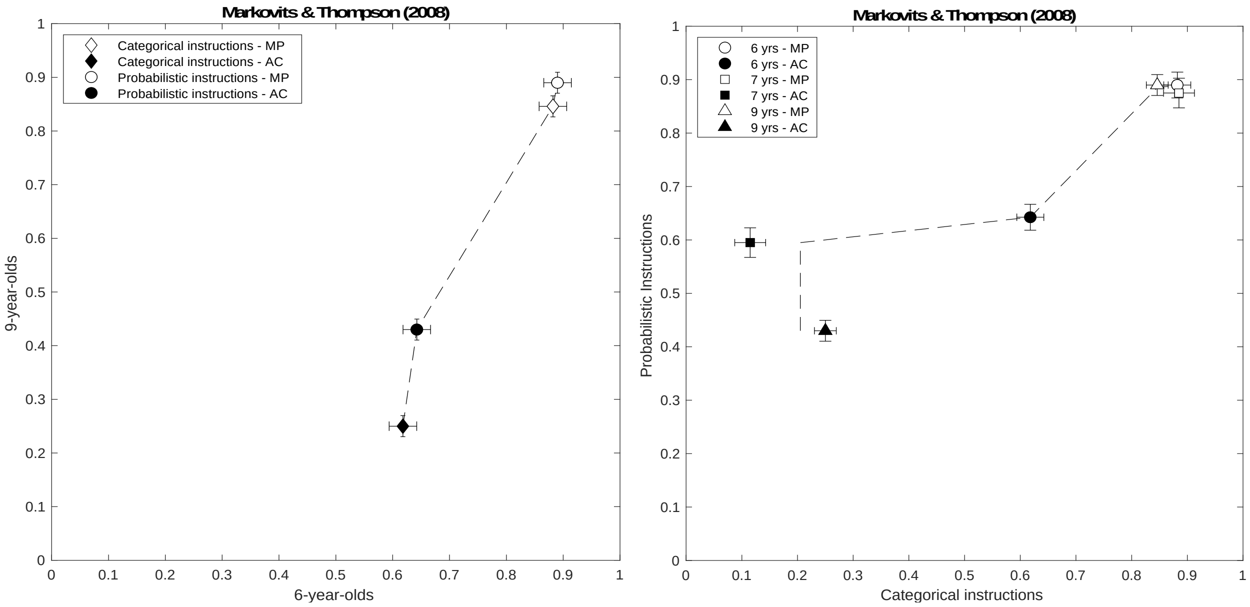

Figure 2. State-trace plots of developmental data from Markovits and Thompson (2008). The dashed line shows the best-fitting monotonic curve. (a) A state-trace plot of data points that are fit perfectly by a monotonic curve. Age groups defined the two dependent variables. Instruction types and logical argument forms defined the four experiment conditions. (b) A state-trace plot with data points showing departures from monotonicity (the two leftmost data points). Instruction types defined the two dependent variables. Age groups and logical argument forms defined the six experiment conditions. (MP = Modus Ponens; AC = Affirmation of the Consequent.)

When state-trace analysis was applied to adult reasoning data, limited support for dual-process accounts was found. Stephens et al. (2019) examined the results from nine papers cited by Evans and Stanovich (2013) as the best experimental evidence for dual processes in adult reasoning. The data were re-analysed using STA to test whether there was sufficient evidence for multiple latent variables. Results showed that the dissociations thought to reflect the operation of separate underlying reasoning processes largely disappear against the stricter criteria of STA. Many of the empirical dissociations were shown to be consistent with one underlying latent reasoning process (Stephens et al., 2019). However, to our knowledge, STA has not yet been applied to developmental reasoning data. Given the observation of complex developmental patterns (e.g., Chiesi et al., 2008; Morsanyi & Handley, 2008), there may be stronger evidence for multiple latent variables in this case.

1.6 Dual processes in reasoning development

While dual-process theories have been developed to explain adult cognition – arguably a period of mature cognitive development (Evans, 2011), this approach has been extended to the important developmental question of how logical reasoning changes from childhood to adulthood.

The key developmental prediction of many dual-process theories has been that analytic processes that can lead to correct logical decisions should become more common with development (e.g., Klaczynski, 2009; Stanovich, & West, 2000). Such a prediction follows from the robust evidence that general intelligence and working memory capacity develop steadily with age and correlate with the controlled Type 2 processing (see Evans, 2011). The predicted developmental outcome has been tested on a wide range of reasoning tasks such as denominator neglect and gambler's fallacy problems in which participants choose between solutions thought to reflect either intuitive Type 1 or reflective Type 2 processing (e.g., Chiesi, Primi, & Morsanyi, 2011; Toplak, West, & Stanovich, 2014). Consistent with the prediction, researchers have reported an age-related increase in normative responses on those tasks (Chiesi et al., 2011; Toplak et al., 2014).

Nonetheless, drawing developmental predictions from dual-process theories has proven to be difficult. Amalgamating research results from different reasoning tasks, there has not been an entirely consistent trend for rational thinking to be monotonically increasing or decreasing with age (Stanovich, Toplak, & West, 2008). Beginning with the findings by Jacobs and Potenza (1991), there is growing evidence for surprising developmental reversals

– negative relations between age and normative responses. For instance, Morsanyi and Handley (2008) observed a marked rise in heuristic responses with age that was related to improvement in cognitive capacity. Adding to the complexity, other studies found that normative responses increased with age on some tasks but decreased with age on other tasks in the same population (e.g., De Neys & Vanderputte, 2011). Moreover, U-shaped patterns of reasoning development have been documented – that normative responses sometimes decline with age before becoming more prevalent again (e.g., Reyna & Farley, 2006).

Dual-process theories have been used in an attempt to account for the complex developmental trajectories in terms of changes in both Type 1 and 2 processing. On the one hand, the age-related increase in heuristic responses in reasoning concerns the developmental course of Type 1 processing and the acquisition of mindware[2] usable by Type 1 processing, as evidence suggests that children rely more and more on relevant knowledge (e.g., social stereotypes) and contextual or pragmatic cues to reason as they get order (Stanovich, West, & Toplak, 2011). On the other hand, Type 2 processing, which is often associated with normative responding, clearly develops with increases in cognitive capacity (Evans, 2011). Importantly, Type 2 processing must be able to suppress Type 1 processing in order to override erroneous automatic outputs (Stanovich et al., 2011). As a result, intuitive reasoning responses cued by context or based on background knowledge may be expected to increase from early to late childhood, and then decline in late adolescence and early adulthood as the ability to interfere with default heuristic responding and override with analytic reasoning finally develops (Evans, 2011). Given the potentially different developmental courses of Type 1 and Type 2 processing and their related mindware, the observed developmental reversals

[2] Mindware is a term that refers to the knowledge, rules, procedures, and strategies that are stored in memory and are available for a person to retrieve when making an inference (Toplak et al., 2014).

and U-shaped patterns on some reasoning tasks thus possibly derive from the complexity of these interacting processes (Stanovich et al., 2011).

In summary, the developmental literature is mixed in its support for a dual-process account of reasoning, which indicates that child development of reasoning is an important context for investigating competing single-process and dual-process theories.

1.7 The current study

In the reasoning literature to date, state-trace analysis has been applied to data from exclusively adult based populations, providing some evidence against the idea that dual-process theories provide a better account than a single underlying reasoning process. However, it remains unclear whether such results hold across the developmental period. To address this research gap, the current study sought to test single- and dual-process models of reasoning in a developmental context. To re-appraise the existing evidence for dual-process models, the goal was to build a developmental database and examine the data using the more rigorous approach of STA. Developmental reasoning studies were searched systematically, and summary statistics (i.e., means, sample size, and standard deviations) were collected from eligible studies. By focusing on developmental studies that have manipulated factors relevant to dual-process theories, this study attempted to maximise the opportunity to detect more than one latent reasoning variable, if this view is correct. The current study is thus well suited to clarify what cognitive mechanism(s) underlie reasoning performance, as well as help guide future experimental tests of competing single-process and dual-process accounts.

1.7.1 Aims

The current study had three specific aims:

1. Build a database of published developmental reasoning studies that have involved a group of children or young adolescents (6-15 years) and to which dual-process accounts have

been applied (i.e., targeting reasoning studies that have manipulated factors relevant to dual-process theories).

2. Re-evaluate the reasoning data using state-trace analysis to determine whether there is sufficient evidence to reject the simplest single-process account with one underlying latent variable, and thus potentially support dual-process theories.

3. Identify any experimental factors or designs that lead to compelling evidence for dual processes in reasoning development, according to the results of the state-trace analyses.

Chapter 2: Methods

2.1 Literature search

This study has adopted a systematic search strategy. An electronic search of the PsycINFO and Scopus databases was undertaken on 14 April 2020 to obtain studies that have applied dual-process theories to examine reasoning in children, up to young adolescents (under 15 years). PsycINFO was chosen as it comprehensively covers scholarly publications in psychology, and the multidisciplinary database Scopus was used for retrieving literature outside the coverage of PsycINFO.

The search strategy involved using the subject heading “reasoning” in combination with key terms related to “dual-process theories” and the target populations. The detailed search terms and limiters are outlined in Figure 3. A preliminary search was conducted to determine the optimal combination of keywords. “Dual strateg*”, a keyword related to dual-process theories, was not used in the final search because it did not contribute to retrieving extra literature. To maximise search results, most keywords were truncated, and the subject heading “reasoning” was exploded[3] when searching PsycINFO on the Ovid retrieval system.

Databases: PsycINFO, Scopus

Search:

reasoning

AND

child* OR adolescen*

AND

“dual process*” OR “two process*” OR “dual system*” OR “two system*”

Figure 3. Search terms and strategy.

[3] The explode function in PsycINFO allows the subject heading and narrower terms below the heading to be searched and thus increases the number of related references being found (Ovid Technologies, 2020).

2.2 Study eligibility

To be eligible for inclusion in the database of this study, all studies published from inception to 14 April 2020 were considered if they met the following inclusion criteria:

1. the study was empirical; and
2. included a sample of children with a mean age between 6 years and 15 years; and
3. used a variant of dual-process theories to account for reasoning responses; and
4. reported summary statistics that were based on non-categorical data; and
5. had sufficient experimental factors and conditions for state-trace analysis (i.e., at least one relevant factor such as a time pressure manipulation that could define the two STA dependent variables and at least three conditions in which these two dependent variables were measured); and
6. was published in a peer-reviewed journal; and
7. had an English-language full-text version available.

The screening process is summarised in Figure 4. A total of 186 studies were identified in the initial database search. Using the reference management software EndNote, all references returned by electronic searches were screened for duplicates. After the removal of duplicates, 140 references were retained. The inclusion criteria were then applied to the titles and abstracts of these references, which resulted in a pool of 106 potentially eligible studies. The full texts for potential studies were subsequently retrieved and examined against the eligibility criteria again to determine suitability for inclusion. The screening was undertaken by the author, with a subset of 20 papers checked by a second researcher. This process resulted in 10 papers providing usable reasoning data for STA.

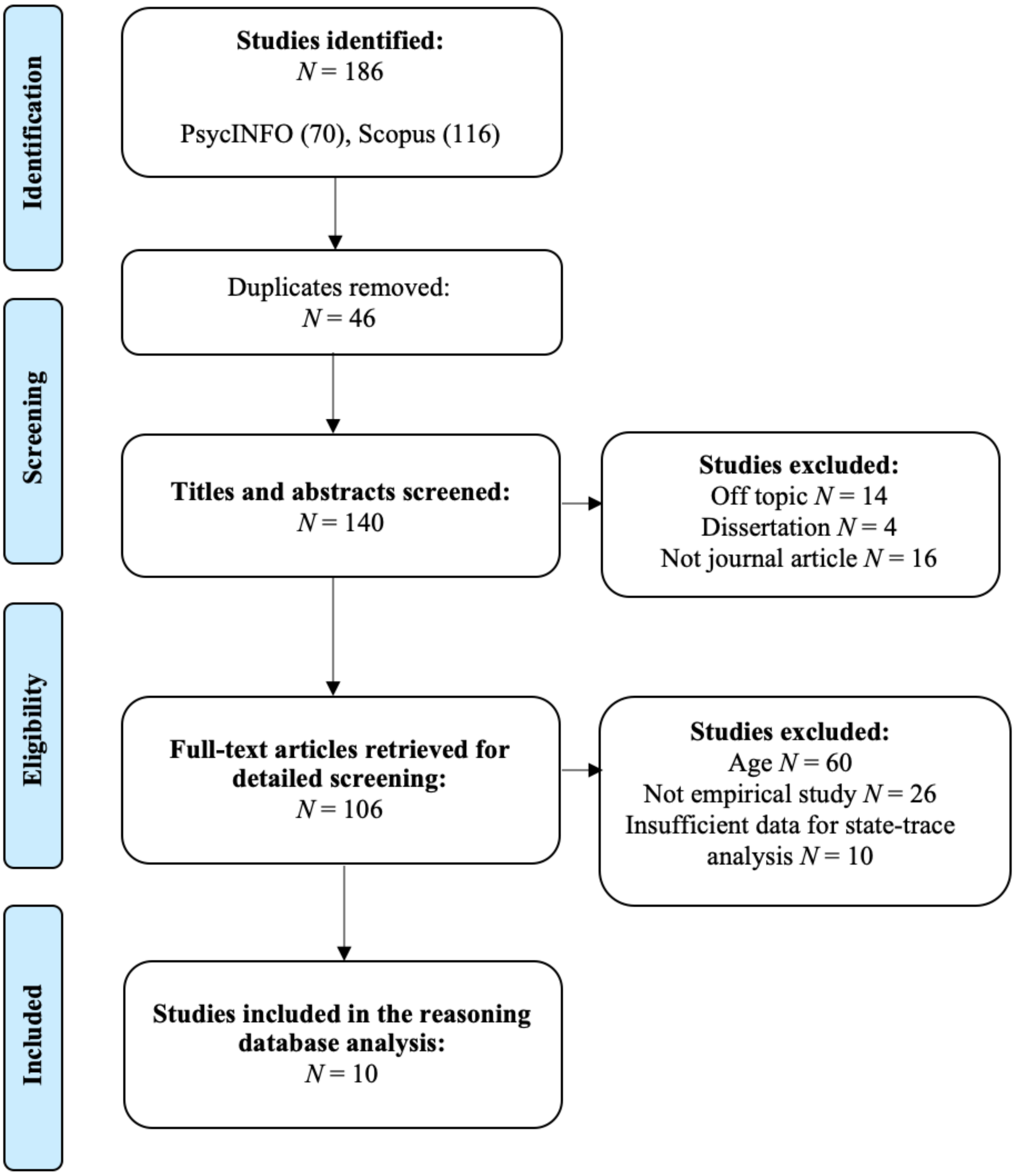

Figure 4. Flowchart of the study selection process.

2.3 Data collection and preparation

The database of this study consisted of 78 datasets constructed from data provided by the 10 papers in the final sample. The data were extracted from experiments that reported condition means, and were based on various types of thinking and reasoning tasks, such as argument evaluation tasks, base-rate neglect problems, contingency table problems, and law of large numbers problems. Details of the experiments are listed in Table 1. The database is available in the supplemental materials[4].

The construction of a dataset for STA requires different treatment of the experiment factors compared to more standard tests such as ANOVA. Given that a state-trace plot reflects the co-variation of two dependent variables across a set of different experimental conditions (Newell & Dunn, 2008), each dataset requires a key experimental factor to define the two dependent variables that will form the two axes of the state-trace plot (e.g., the factor categorical/probabilistic instructions defined the two dependent variables in Figure 2b). Crucially, to test dual-process models of reasoning, each dependent variable should plausibly be differentially influenced by Type 1 and Type 2 processing (Stephens et al., 2019). Other factors in the dataset should then be treated as independent variables that define the experimental conditions across the dependent variables, that is, data points in the state-trace plot (e.g., the factors, age and logical forms, defined the six conditions in Figure 2b).

Accordingly, for the current datasets, the factor selected to define the dependent variables was the factor highlighted by experimenters and/or dual-process theories to have levels that differentially reflect the two types of processing. One type of such factor concerned the task characteristics. That is, the reasoning task itself (e.g., base-rate neglect problems; Felmban & Klaczynksi, 2019) contained conflicting information in which one kind of information was thought to cue Type 1 processing (e.g., stereotype-consistent evidence) and the other kind of information was thought to cue Type 2 processing (e.g., base rate evidence). Another type of dependent variable concerned additional experimental manipulations that were thought to enhance or inhibit Type 2 processing, such as un-speeded/speeded judgements (Markovits et al., 2019) and categorical/probabilistic

[4] The database can also be accessed via the link https://universityofadelaide.box.com/s/w3y8t2x5d626twesaxb181odz5amrwmi.

instructions (Markovits & Thompson, 2008). Lastly, where possible, different age groups were also used to define the two dependent variables. Age was a sensible factor in that dual-process theories are fairly consistent in their predictions that reflective Type 2 processes develop with age as cognitive capacity increases (e.g., Evans, 2011; Stanovich & West, 2000). Also, the individual differences literature in child development supports the idea that increasing chronological age is accompanied with improved cognitive abilities (e.g., Nettelbeck & Burns, 2010). Thus, responses from an older age group may reflect more Type 2 processing than from a younger age group. Furthermore, when multiple factors within an experiment were available to define the two dependent variables, different datasets were constructed for the same experiment to allow each factor to define the dependent variables in turn. In this way, the database maximised the opportunity for finding evidence against one underlying latent variable. The alternative datasets were signalled as "factors inverted" in the database (e.g., dataset 62, see section 3.2 for an illustrative example).

Summary statistics for each experimental condition were extracted from the relevant sections in each publication, when available. These statistics included the means, sample size (*N*s), and standard deviations (*SD*s). Plot digitizing software was used to extract data from figures. Where only standard errors (*SE*s) were reported, *SE*s were converted to *SD*s by the author. As a range of behaviour measures were used for the dependent variables (e.g., percentages of normative responses, ratings for conclusion probability), scores were set to a 0-1 proportion scale for consistency across datasets from different experiments. Datasets for Lin and Shih (2016) were an exception. Original scores reported by the authors were retained in the database because scores across the thinking tasks were already standardised and the measurement schemes were not clearly detailed in their paper.

2.4 State-trace analysis tools and application

Data were analysed using MATLAB (Version R2020a). For each dataset, a state-trace plot was produced and the conjoint monotonic regression (CMR) test (Dunn & Kalish, 2018; Kalish, Dunn, Burdakov, & Sysoev, 2016) was applied to test whether there was statistically significant evidence against a single underlying latent variable. In general, a state-trace plot is produced by plotting the mean of each experimental condition for one dependent variable against the corresponding mean for the other dependent variable. Data points falling on a single monotonic curve correspond to a "one-dimensional" state-trace, indicating that the observed data are consistent with one latent variable. Departures from monotonicity would instead imply the operation of more than one latent variable. For instance, such a "two-dimensional" state-trace would appear when data points form two separate monotonic curves, or a "cloud" of data points, suggesting the operation of multiple latent variables.

In a similar vein, the CMR test examined the experimental conditions across the two dependent variables of each dataset and found a best-fitting (increasing) monotonic approximation of the observed data, via a custom optimization algorithm (Dunn & Kalish, 2018; Kalish et al., 2016). A formal statistical test of the hypothesis that the monotonic approximation provides an adequate fit to the data was then conducted. In this test, the observed fit of the single-latent-variable model was compared to a bootstrap sample distribution of fit values under the hypothesis that the model was true, and a *p*-value was obtained. For all CMR tests, 10,000 bootstrap samples were used (Type 1 error rate, $\alpha = .05$).

The means in each dataset were used by the CMR algorithm to search for the best-fitting monotonic function and the *SD*s were used in the bootstrapping procedure for estimating the variability of the data. As the database contained summary statistics only, a parametric bootstrap distribution was necessary in that the observed data for each experimental condition were assumed to be normally distributed. However, *SD*s or *SE*s for relevant experimental conditions were not reported in experiments conducted by Ameel, Verschueren, and Schaeken (2007) or Markovits and Thompson (2008). Therefore, for these studies, the CMR tests were performed on estimated *SD*s. The three levels of possible *SD*s used in the Stephens et al. (2019) state-trace analysis of adult reasoning data were adopted: .10, .20, and .30, with .10 being optimistic – note that the mean for known *SD*s in the current database (based on the 0-1 scale) was .26. For nine Markovits and Thompson datasets (datasets 69-77), the raw data were actually dichotomous in nature as there was only one trial per cell, per participant; however, the group means could still be examined paired with estimated *SD*s of .10, .20, and .30, to simulate potential evidence, had more data been collected. Last, approximate cell *N*s were used for each condition of Hagá, Garcia-Marques, and Olson (2014). As only the total experiment *N*s were reported, the even allocation of *N*s to conditions was assumed.

Chapter 3: Results

3.1 State-trace analysis

The CMR tests were applied to 62 datasets with known *SD*s and 16 datasets with unknown *SD*s. The CMR test results for the total of 78 datasets are presented in Table 1, and sample state-trace plots with the best fitting monotonic curves are illustrated in Section 3.2 (State-trace plots for all datasets are included in the supplemental materials[5]).

Overall, 19 out of the 78 datasets (24%) returned a significant result on the CMR tests ($p < .05$) when the most realistic *SD*s of .30 were assumed for datasets with unknown *SD*s. If *SD*s of .10 or .20 were instead assumed for those datasets, 29 out of the 78 tests (37%) or 24 out of the 78 tests (31%) were statistically significant ($p < .05$) respectively.

After the example datasets in the next section, the results were further considered in two groups, based on the type of factor that defined the two dependent variables in state-trace analysis: first, dependent variables based on experimental manipulations, which included factors concerning task characteristics (e.g., familiar versus unfamiliar stereotype cues in base-rate tasks) or additional experimental manipulations (e.g., un-speeded versus speeded conditions); and second, dependent variables based on age groups.

3.2 Illustrative examples

Figures 5-8 show four examples of state-trace plots from the database: two significant (Figures 5 and 6) and two non-significant (Figures 7 and 8) instances. The corresponding datasets for the four sample plots are datasets 60, 61, 6, and 62 from the database.

The dashed line superimposed on the data points of each figure corresponded to the predictions of the best-fitting monotonic model, found by the CMR algorithm. It was apparent from both Figures 5 and 6 that the monotonic model fitted the data poorly. In Figure

---

[5] The plots can also be viewed via the link https://universityofadelaide.box.com/s/w3y8t2x5d626twesaxb181odz5amrwmi.

5, the data points in the state-trace substantially deviated from the dashed line, which strongly suggested the state-trace was two-dimensional – or at least not monotonically *increasing* (there was actually a negative relationship between conditions in this case). Dataset 60 further returned a significant result on the CMR test ($p < .001$), indicating that these data were consistent with an underlying model with more than one latent variable. The departure of two points from the dashed line in dataset 61(Figure 6) similarly suggested the state-trace was not one-dimensional, and the CMR test for this dataset was statistically significant ($p = .04$).

The visual inspection of Figures 7 and 8 suggested that each monotonic curve found by the CMR procedure fitted the data very well, as all data points fell along (see Figure 7) or close to (see Figure 8) a monotonically increasingly curve, leading to the inference that there was little evidence to reject the null hypothesis of a single latent variable. Correspondingly, neither dataset 6 ($p = 1.00$; Figure 7) nor dataset 62 ($p = .75$; Figure 8) returned a statistically significant result on the CMR test. These results indicated that the data were consistent with a model in which the effects of the independent variables on the two dependent variables were mediated by one latent variable.

Moreover, note that Figure 8 is an example of rearranging a single dataset (cf. Figure 6). Datasets 61 and 62 have both used the data from Markovits et al. (2019). The time pressure factor and the age factor defined the two dependent variables for datasets 61 and 62 respectively. Notably, dataset 61 returned a significant result on the CMR test, whereas dataset 62 did not. Thus, the different CMR test results implied that time pressure was a more powerful factor than age in leading to potential evidence for dual processes in reasoning.

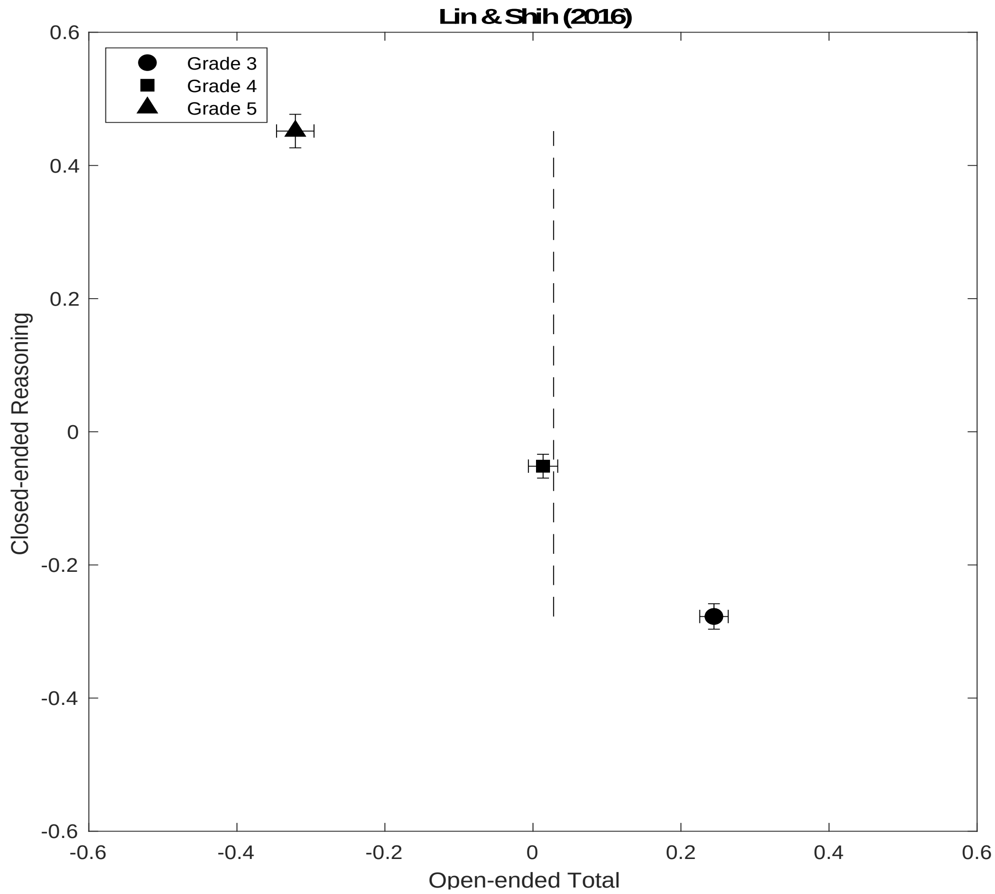

Figure 5. State-trace plot for dataset 60 in the reasoning database, from Lin and Shih (2016). Task characteristics (i.e., open-ended versus closed-ended thinking) defined the two dependent variables. Age groups defined the three experiment conditions. Reasoning performance was measured using the standardised scores for the thinking tasks. The dashed-line shows the best-fitting monotonic curve. Error bars show *SE*s of the mean of each experiment condition.

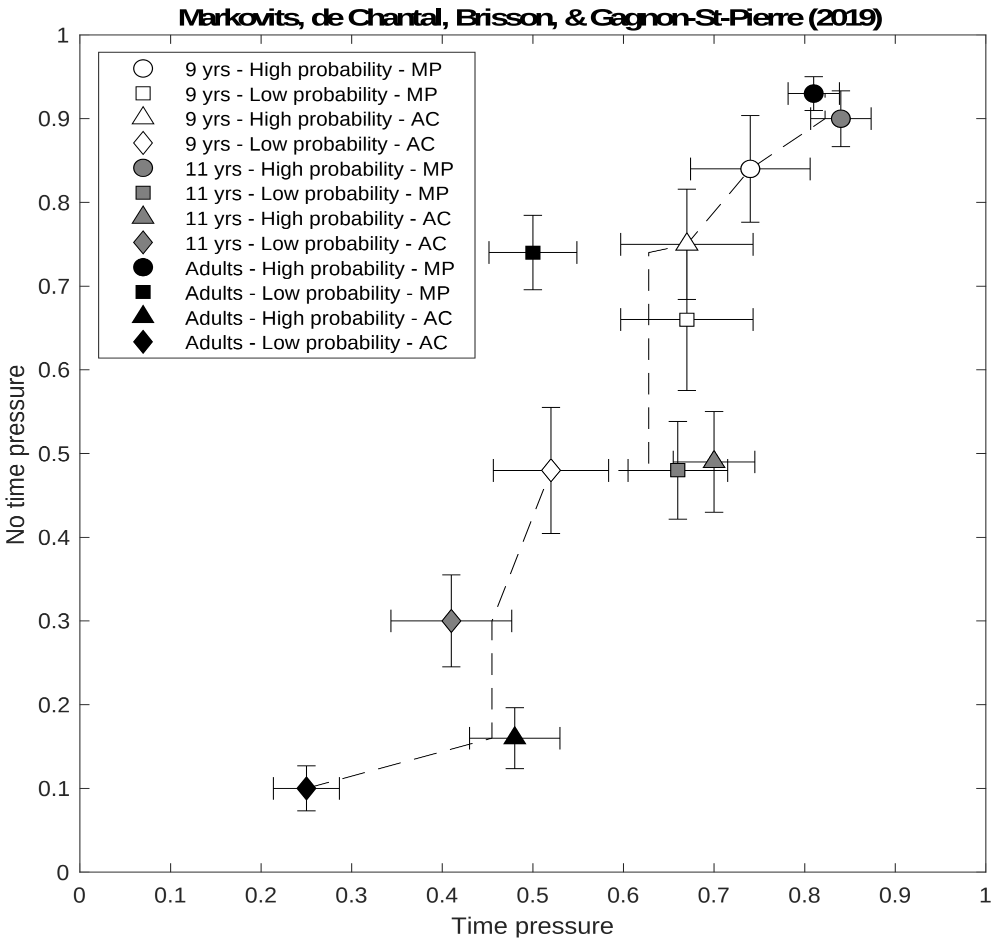

Figure 6. State-trace plot for dataset 61 in the reasoning database, from Markovits et al. (2019). Time pressure defined the two dependent variables. Age groups, conclusion probability, and logical argument forms (MP versus AC) defined the twelve experiment conditions. Reasoning performance was measured using proportion of conclusion acceptance in an argument evaluation task. The dashed-line shows the best-fitting monotonic curve. Error bars show *SE*s of the mean of each experiment condition.

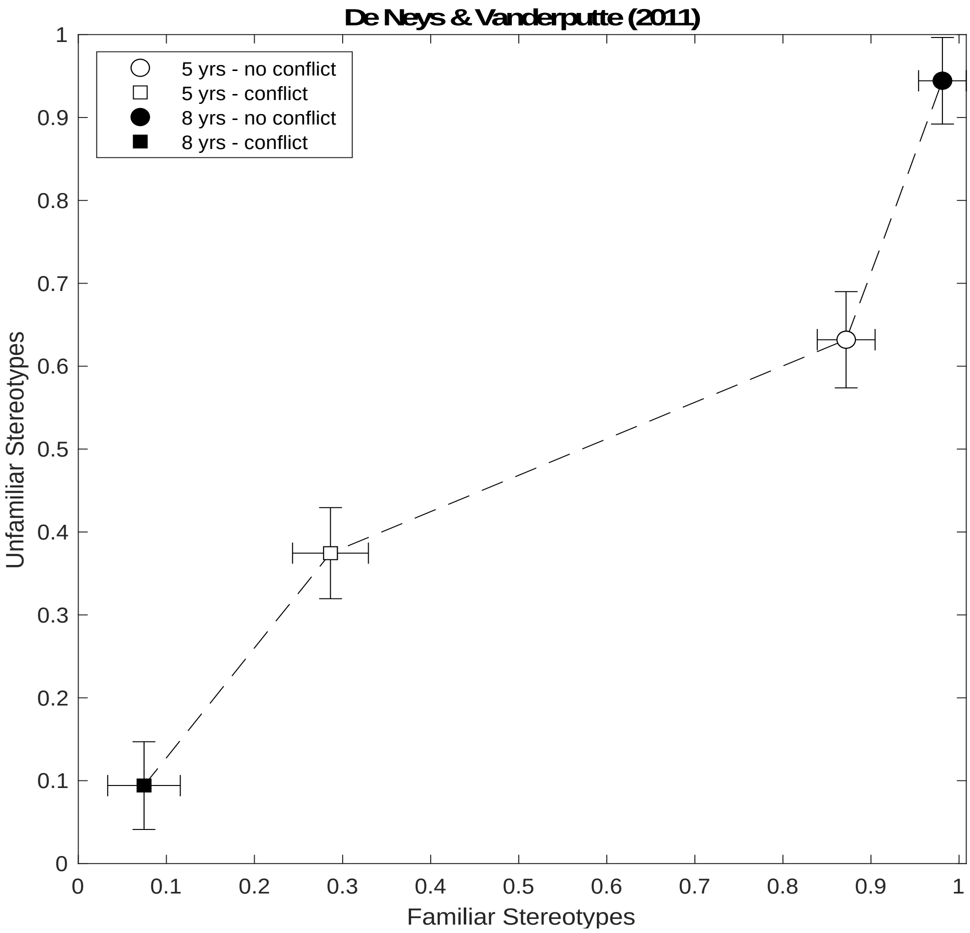

Figure 7. State-trace plot for dataset 6 in the reasoning database, from De Neys and Vanderputte (2011). Stereotype familiarity defined the two dependent variables. Age groups and conflict defined the four experiment conditions. Reasoning performance was measured as the percentage of correct responses in base-rate tasks. The dashed-line shows the best-fitting monotonic curve. Error bars show *SE*s of the mean of each experiment condition.

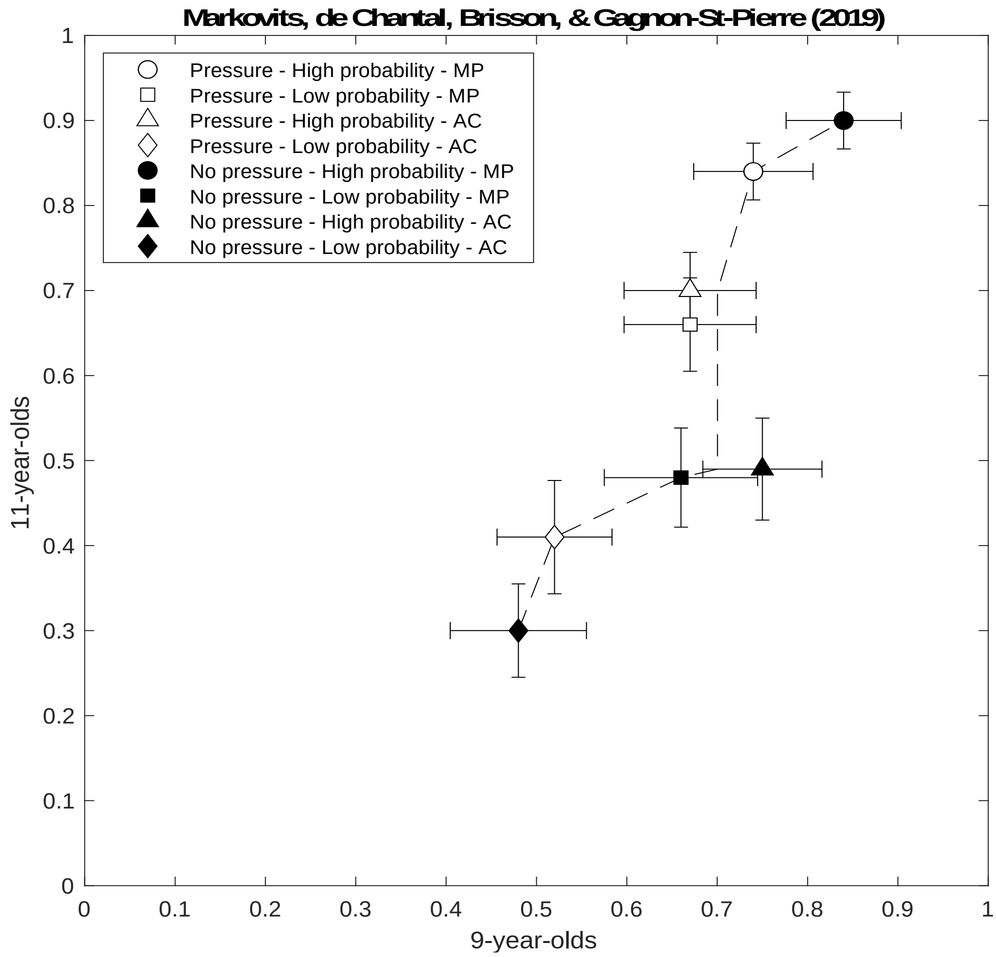

Figure 8. State-trace plot for dataset 62 in the reasoning database, from Markovits et al. (2019). Age groups defined the two dependent variables. Time pressure, conclusion probability, and logical argument forms (MP versus AC) defined the eight experiment conditions. Reasoning performance was measured using proportion of conclusion acceptance in an argument evaluation task. The dashed-line shows the best-fitting monotonic curve. Error bars show *SE*s of the mean of each experiment condition.

3.3 Dependent variables based on experimental manipulations

In the first group of 44 datasets where the factor selected to define the STA dependent variables was based on experimental manipulations, the CMR tests were applied to 38 datasets with reported *SD*s from the original experiments and 6 datasets with estimated *SD*s of .10, .20, and .30. For datasets with known *SD*s, the null hypothesis of monotonicity was retained for 21 (55%) datasets ($p > .05$). These non-significant results indicated that the best fitting monotonic curve found by the CMR procedure for each dataset provided an adequate fit to the data and were consistent with one latent variable. For the other 17 (45%) datasets with reported *SD*s, the results of the CMR tests were statistically significant ($p < .05$), suggesting that the observed data were consistent with an underlying model with more than one latent variable or psychological process. The null results were from De Neys and Vanderputte (2011) and Hagá et al. (2014), whereas the significant results included one Markovits et al. (2019) dataset (dataset 61). While the Klaczynski (2000) data and Lin and Shih (2016) data yielded mixed results, the former had only one dataset that returned a significant result, but the latter contributed to 15 positive results.

As for the 6 datasets with unknown *SD*s, if low *SD*s of .10 were assumed for these datasets, 5 out of the 6 CMR tests (83%) were statistically significant ($p < .05$), suggesting that more than one latent variable was required to explain the data. However, if *SD*s of .20 or .30 were assumed for these datasets, the number of significant results returned by the CMR tests dropped to 3 (50%) or 1 (17%) respectively, indicating the data were either ambivalent or more consistent with only one latent variable. For instance, regarding dataset 1 from Ameel et al. (2007) in which the task characteristics (i.e., the white-block problems were assumed to inhibit Type 1 processing) defined the STA dependent variables, the null hypothesis of a single latent variable could only be rejected when the estimated *SD*s of .10 was applied in the CMR test ($p = .02$), whereas the null hypothesis was retained when the more probable *SD*s of .20 and .30 were used, $p = .21$ and $p = .38$, respectively.

3.4 Dependent variables based on age groups

With respect to the 34 datasets in which age was the factor chosen to define the STA dependent variables, the CMR tests were performed on 24 datasets with reported *SD*s and 10 datasets with three levels of plausible, estimated *SD*s: .10, .20, and .30. For datasets with known *SD*s, 23 out of the 24 datasets (96%) returned a result that was not statistically significant on the CMR test ($p > .05$), failing to detect evidence for more than one latent variable. The null results included data from De Neys and Vanderputte (2011), Felmban and Klaczynski (2019), Kail (2013), Klaczynski (2000), Markovits and Thompson (2008), and Obersteiner, Bernhard, and Reiss (2015). While some datasets from Markovits et al. (2019) have returned non-significant results, the only dataset (dataset 64) found to be statistically significant was also from Markovits et al. ($p = .01$), involving 11-year-old children and adults evaluating conclusions of conditional arguments, when given limited versus unlimited time.

As with the previous group of datasets with assumed *SD*s, positive CMR results for the 10 datasets with unknown *SD*s gradually diminished when the estimated *SD*s changed from a low .10 to a moderate .30. When *SD*s of .10 or .20 were assumed for the datasets, the CMR tests respectively returned 6 and 3 statistically significant results ($p < .05$), providing potential evidence for more than one latent variable. However, if more realistic *SD*s of .30 were assumed for these datasets, none of the CMR tests returned a positive result, suggesting that there was no evidence for more than one latent variable in the data.

Table 1

Details of the reasoning database and the state-trace analysis results (based on reported or estimated SDs).

	Source	Task	Dependent variable 1	Dependent variable 2	Factors that define the conditions	*p-values*			
						*SD*s reported	*SD*s = .10	*SD*s = .20	*SD*s = .30
1	Ameel, Verschueren, & Schaeken (2007) Experiments 1 & 2	Transitive reasoning task	Proportion correct – White-block problems	Proportion correct – Coloured-block problems	1) 8 yrs / 9 yrs 2) Experiment 1 (No context) / Experiment 2 (Context)	–	**.015**	.207	.384
2	Ameel, Verschueren, & Schaeken (2007) Experiments 1 & 2 [factors inverted]	Transitive reasoning task	Proportion correct – Experiment 1 (No context)	Proportion correct – Experiment 2 (Context)	1) 8 yrs / 9 yrs 2) White-block / coloured-block problems	–	.566	.724	.803
3	Ameel, Verschueren, & Schaeken (2007) Experiments 1 & 2 [factors inverted]	Transitive reasoning task	Proportion correct – 8-year-olds	Proportion correct – 9-year-olds	1) Experiment 1 (No context) / Experiment 2 (Context) 2) White-block / coloured-block problems	–	**.019**	.225	.416
4	De Neys & Vanderputte (2011)	Base-rate problems [Table 1]	Percent accuracy – Familiar stereotypes	Percent accuracy – Unfamiliar stereotypes	1) 5 yrs / 8 yrs 2) Conflict / no conflict	.129	–	–	–
5	De Neys & Vanderputte (2011) [factors inverted]	Base-rate problems [Table 1]	Percent accuracy – 5-year-olds	Percent accuracy – 8-year-olds	1) Familiar / unfamiliar stereotypes 2) Conflict / no conflict	.161	–	–	–

6	De Neys & Vanderputte (2011)	Base-rate problems [Figure 4]	Percent accuracy – Familiar stereotypes	Percent accuracy – Unfamiliar stereotypes	1) 5 yrs / 8 yrs 2) Conflict / no conflict	1.000	–	–	–
7	De Neys & Vanderputte (2011) [factors inverted]	Base-rate problems [Figure 4]	Percent accuracy – 5-year-olds	Percent accuracy – 8-year-olds	1) Familiar / unfamiliar stereotypes 2) Conflict / no conflict	1.000	–	–	–
8	Felmban & Klaczynski (2019)	Base-rate problems [Table 1]	Proportion correct – 10-year-olds	Proportion correct – 13-year-olds	1) Stereotypical / anecdotal problems 2) Conflict / no conflict	.334	–	–	–
9	Felmban & Klaczynski (2019)	Base-rate problems [Table 1]	Proportion correct – 10-year-olds	Proportion correct – 16-year-olds	1) Stereotypical / anecdotal problems 2) Conflict / no conflict	.357	–	–	–
10	Felmban & Klaczynski (2019)	Base-rate problems [Table 1]	Proportion correct – 13-year-olds	Proportion correct – 16-year-olds	1) Stereotypical / anecdotal problems 2) Conflict / no conflict	1.000	–	–	–
11	Felmban & Klaczynski (2019)	Base-rate problems [Table 2]	Possibility ratings – 10-year-olds	Possibility ratings – 13-year-olds	1) Stereotypical / anecdotal problems 2) Conflict / no conflict	.384	–	–	–
12	Felmban & Klaczynski (2019)	Base-rate problems [Table 2]	Possibility ratings – 10-year-olds	Possibility ratings – 16-year-olds	1) Stereotypical / anecdotal problems 2) Conflict / no conflict	.405	–	–	–
13	Felmban & Klaczynski (2019)	Base-rate problems [Table 2]	Possibility ratings – 13-year-olds	Possibility ratings – 16-year-olds	1) Stereotypical / anecdotal problems 2) Conflict / no conflict	1.000	–	–	–
14	Felmban & Klaczynski (2019)	Base-rate problems [Table C1]	Proportion correct – 10-year-olds	Proportion correct – 13-year-olds	1) Gender / obesity problems 2) Conflict / no conflict	.300	–	–	–

15	Felmban & Klaczynski (2019)	Base-rate problems [Table C1]	Proportion correct – 10-year-olds	Proportion correct – 16-year-olds	1) Gender / obesity problems 2) Conflict / no conflict	.286	–	–	–
16	Felmban & Klaczynski (2019)	Base-rate problems [Table C1]	Proportion correct – 13-year-olds	Proportion correct – 16-year-olds	1) Gender / obesity problems 2) Conflict / no conflict	.298	–	–	–
17	Hagá, Garcia-Marques, & Olson (2014) Study 1	Social inference problems	Dispositional ratings – Gift condition	Dispositional ratings – Punishment condition	1) Kindergarteners / second graders / sixth graders / ninth graders / undergraduates	.622	–	–	–
18	Hagá, Garcia-Marques, & Olson (2014) Study 2	Social inference problems	Situational ratings – Cheerful condition	Situational ratings – Cry-baby condition	1) Kindergarteners / second graders / sixth graders / ninth graders / undergraduates	.313	–	–	–
19	Hagá, Garcia-Marques, & Olson (2014) Studies 3a & 3b	Social inference problems	Dispositional ratings – Choice condition	Dispositional ratings – No-choice condition	1) Kindergarteners / second graders / sixth graders / ninth graders / undergraduates / adults 2) Study 3a / 3b	.970	–	–	–
20	Kail (2013) Experiment 1 & 2	Vignettes with statistical vs. testimonial evidence	Confidence ratings – 9-year-olds	Confidence ratings – 13-year-olds	1) Casual / expert testimony, weak / strong statistical evidence	.156	–	–	–
21	Klaczynski (2000)	Experiment evaluation problems	Strength scores – Early adolescents	Strength scores – Middle adolescents	1) Favourable / neutral / unfavourable 2) Social class / religion	.055	–	–	–
22	Klaczynski (2000)	Experiment evaluation problems	Validity scores – Early adolescents	Validity scores – Middle adolescents	1) Favourable / neutral / unfavourable 2) Social class / religion	.126	–	–	–

23	Klaczynski (2000)	Experiment evaluation problems	Justification scores – Early adolescents	Justification scores – Middle adolescents	1) Favourable / neutral / unfavourable 2) Social class / religion	.838	–	–	–
24	Klaczynski (2000)	Experiment evaluation problems	Implausibility scores – Early adolescents	Implausibility scores – Middle adolescents	1) Favourable / neutral / unfavourable 2) Social class / religion	.637	–	–	–
25	Klaczynski (2000)	Law of large numbers problems	Strength scores – Early adolescents	Strength scores – Middle adolescents	1) Favourable / neutral / unfavourable 2) Social class / religion	.511	–	–	–
26	Klaczynski (2000)	Law of large numbers problems	Persuasiveness scores – Early adolescents	Persuasiveness scores – Middle adolescents	1) Favourable / neutral / unfavourable 2) Social class / religion	.720	–	–	–
27	Klaczynski (2000)	Law of large numbers problems	Justification scores – Early adolescents	Justification scores – Middle adolescents	1) Favourable / neutral / unfavourable 2) Social class / religion	.863	–	–	–
28	Klaczynski (2000)	Law of large numbers problems	Implausibility scores – Early adolescents	Implausibility scores – Middle adolescents	1) Favourable / neutral / unfavourable 2) Social class / religion	.720	–	–	–
29	Klaczynski (2000) [factors inverted]	Experiment evaluation problems	Strength scores – Favourable	Strength scores – Unfavourable	1) Early / middle adolescents 2) Social class / religion	.495	–	–	–
30	Klaczynski (2000) [factors inverted]	Experiment evaluation problems	Validity scores – Favourable	Validity scores – Unfavourable	1) Early / middle adolescents 2) Social class / religion	.305	–	–	–
31	Klaczynski (2000) [factors inverted]	Experiment evaluation problems	Justification scores – Favourable	Justification scores – Unfavourable	1) Early / middle adolescents 2) Social class / religion	.465	–	–	–
32	Klaczynski (2000) [factors inverted]	Experiment evaluation problems	Implausibility scores – Favourable	Implausibility scores – Unfavourable	1) Early / middle adolescents 2) Social class / religion	**.013**	–	–	–

33	Klaczynski (2000) [factors inverted]	Law of large numbers problems	Strength scores – Favourable	Strength scores – Unfavourable	1) Early / middle adolescents 2) Social class / religion	.485	–	–	–
34	Klaczynski (2000) [factors inverted]	Law of large numbers problems	Persuasiveness scores – Favourable	Persuasiveness scores – Unfavourable	1) Early / middle adolescents 2) Social class / religion	.594	–	–	–
35	Klaczynski (2000) [factors inverted]	Law of large numbers problems	Justification scores – Favourable	Justification scores – Unfavourable	1) Early / middle adolescents 2) Social class / religion	.428	–	–	–
36	Klaczynski (2000) [factors inverted]	Law of large numbers problems	Implausibility scores – Favourable	Implausibility scores – Unfavourable	1) Early / middle adolescents 2) Social class / religion	.117	–	–	–
37	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended fluency	Standardised scores – Closed-ended CWRAT	1) Grade 3 / 4 / 5	.331	–	–	–
38	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended fluency	Standardised scores – Closed-ended insight	1) Grade 3 / 4 / 5	.281	–	–	–
39	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended fluency	Standardised scores – Closed-ended reasoning	1) Grade 3 / 4 / 5	.321	–	–	–
40	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended flexibility	Standardised scores – Closed-ended CWRAT	1) Grade 3 / 4 / 5	.607	–	–	–

41	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended flexibility	Standardised scores – Closed-ended insight	1) Grade 3 / 4 / 5	.579	–	–	–
42	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended flexibility	Standardised scores – Closed-ended reasoning	1) Grade 3 / 4 / 5	.602	–	–	–
43	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended originality	Standardised scores – Closed-ended CWRAT	1) Grade 3 / 4 / 5	**.014**	–	–	–
44	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended originality	Standardised scores – Closed-ended insight	1) Grade 3 / 4 / 5	**.015**	–	–	–
45	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended originality	Standardised scores – Closed-ended reasoning	1) Grade 3 / 4 / 5	**.015**	–	–	–
46	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended total	Standardised scores – Closed-ended CWRAT	1) Grade 3 / 4 / 5	.232	–	–	–
47	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended total	Standardised scores – Closed-ended insight	1) Grade 3 / 4 / 5	.223	–	–	–
48	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended total	Standardised scores – Closed-ended reasoning	1) Grade 3 / 4 / 5	.230	–	–	–

49	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Figure 1]	Standardised scores – Open-ended fluency	Standardised scores – Closed-ended CWRAT	1) Grade 3 / 4 / 5	**<.001**	–	–	–
50	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Figure 1]	Standardised scores – Open-ended fluency	Standardised scores – Closed-ended insight	1) Grade 3 / 4 / 5	**<.001**	–	–	–
51	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Figure 1]	Standardised scores – Open-ended fluency	Standardised scores – Closed-ended reasoning	1) Grade 3 / 4 / 5	**<.001**	–	–	–
52	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Figure 1]	Standardised scores – Open-ended flexibility	Standardised scores – Closed-ended CWRAT	1) Grade 3 / 4 / 5	**<.001**	–	–	–
53	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Figure 1]	Standardised scores – Open-ended flexibility	Standardised scores – Closed-ended insight	1) Grade 3 / 4 / 5	**<.001**	–	–	–
54	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Figure 1]	Standardised scores – Open-ended flexibility	Standardised scores – Closed-ended reasoning	1) Grade 3 / 4 / 5	**<.001**	–	–	–
55	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Figure 1]	Standardised scores – Open-ended originality	Standardised scores – Closed-ended CWRAT	1) Grade 3 / 4 / 5	**<.001**	–	–	–
56	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Figure 1]	Standardised scores – Open-ended originality	Standardised scores – Closed-ended insight	1) Grade 3 / 4 / 5	**<.001**	–	–	–

57	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Figure 1]	Standardised scores – Open-ended originality	Standardised scores – Closed-ended reasoning	1) Grade 3 / 4 / 5	**<.001**	–	–	–
58	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Figure 1]	Standardised scores – Open-ended total	Standardised scores – Closed-ended CWRAT	1) Grade 3 / 4 / 5	**<.001**	–	–	–
59	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended total	Standardised scores – Closed-ended insight	1) Grade 3 / 4 / 5	**<.001**	–	–	–
60	Lin & Shih (2016)	Open-ended & closed-ended thinking tasks [Table 3]	Standardised scores – Open-ended total	Standardised scores – Closed-ended reasoning	1) Grade 3 / 4 / 5	**<.001**	–	–	–
61	Markovits, de Chantal, Brisson, & Gagnon-St-Pierre (2019) Studies 1a & 2	Argument evaluation tasks	Proportion correct – Time pressure	Proportion correct – No time pressure	1) 9 yrs / 11 yrs / adults 2) High / low probability 3) MP / AC	**.036**	–	–	–
62	Markovits, de Chantal, Brisson, & Gagnon-St-Pierre (2019) Studies 1a & 2 [factors inverted]	Argument evaluation tasks	Proportion correct – 9-year-olds	Proportion correct – 11-year-olds	1) Time pressure / no time pressure 2) High / low probability 3) MP / AC	.753	–	–	–
63	Markovits, de Chantal, Brisson, & Gagnon-St-Pierre (2019) Studies 1a & 2 [factors inverted]	Argument evaluation tasks	Proportion correct – 9-year-olds	Proportion correct – Adults	1) Time pressure / no time pressure 2) High / low probability 3) MP / AC	.279	–	–	–

64	Markovits, de Chantal, Brisson, & Gagnon-St-Pierre (2019) Studies 1a & 2 [factors inverted]	Argument evaluation tasks	Proportion correct – 11-year-olds	Proportion correct – Adults	1) Time pressure / no time pressure 2) High / low probability 3) MP / AC	**.008**	–	–	–
65	Markovits & Thompson (2008) Study 1	Argument evaluation tasks [Table 1]	Percentages of acceptance – Categorical instructions	Probability ratings – Probabilistic instructions	1) 6 yrs / 7 yrs / 9 yrs 2) MP / AC	–	**.001**	.078	.272
66	Markovits & Thompson (2008) Study 1 [factors inverted]	Argument evaluation tasks [Table 1]	Percentages of acceptance / probability ratings – 6-year-olds	Percentages of acceptance / probability ratings – 7-year-olds	1) Categorical / probabilistic instructions 2) MP / AC	–	.461	.562	.655
67	Markovits & Thompson (2008) Study 1 [factors inverted]	Argument evaluation tasks [Table 1]	Percentages of acceptance / probability ratings – 6-year-olds	Percentages of acceptance / probability ratings – 9-year-olds	1) Categorical / probabilistic instructions 2) MP / AC	–	1.000	1.000	1.000
68	Markovits & Thompson (2008) Study 1 [factors inverted]	Argument evaluation tasks [Table 1]	Percentages of acceptance / probability ratings – 7-year-olds	Percentages of acceptance / probability ratings – 9-year-olds	1) Categorical / probabilistic instructions 2) MP / AC	–	.318	.373	.453
69	Markovits & Thompson (2008) Study 1	Argument evaluation tasks [Table 3]	Percentages of acceptance – Categorical instructions	Probability ratings – Probabilistic instructions	1) 6 yrs / 7 yrs / 9 yrs 2) Alternative produced / no alternative	–	**<.001**	**.005**	.073
70	Markovits & Thompson (2008) Study 1	Argument evaluation tasks [Table 3]	Percentages of acceptance / probability ratings – Alternative produced	Percentages of acceptance / probability ratings – No alternative	1) 6 yrs / 7 yrs / 9 yrs 2) Categorical / probabilistic instructions	–	**<.001**	**<.001**	**.001**

71	Markovits & Thompson (2008) Study 1 [factors inverted]	Argument evaluation tasks [Table 3]	Percentages of acceptance / probability ratings – 6-year-olds	Percentages of acceptance / probability ratings – 7-year-olds	1) Categorical / probabilistic instructions 2) Alternative produced / no alternative	–	**.002**	.088	.222
72	Markovits & Thompson (2008) Study 1 [factors inverted]	Argument evaluation tasks [Table 3]	Percentages of acceptance / probability ratings – 6-year-olds	Percentages of acceptance / probability ratings – 9-year-olds	1) Categorical / probabilistic instructions 2) Alternative produced / no alternative	–	**<.001**	**.044**	.151
73	Markovits & Thompson (2008) Study 1 [factors inverted]	Argument evaluation tasks [Table 3]	Percentages of acceptance / probability ratings – 7-year-olds	Percentages of acceptance / probability ratings – 9-year-olds	1) Categorical / probabilistic instructions 2) Alternative produced / no alternative	–	**<.001**	**.005**	.054
74	Markovits & Thompson (2008) Study 2	Argument evaluation tasks	Percentages of acceptance – Categorical instructions	Probability ratings – Probabilistic instructions	1) 6 yrs / 7 yrs / 9 yrs 2) High / low probability 3) Deductive first / probabilistic first	–	**<.001**	**.042**	.365
75	Markovits & Thompson (2008) Study 2 [factors inverted]	Argument evaluation tasks	Percentages of acceptance / probability ratings – 6-year-olds	Percentages of acceptance / probability ratings – 7-year-olds	1) Categorical / probabilistic instructions 2) High / low probability 3) Deductive first / probabilistic first	–	**<.001**	**.022**	.186
76	Markovits & Thompson (2008) Study 2 [factors inverted]	Argument evaluation tasks	Percentages of acceptance / probability ratings – 6-year-olds	Percentages of acceptance / probability ratings – 9-year-olds	1) Categorical / probabilistic instructions 2) High / low probability 3) Deductive first / probabilistic first	–	.178	.689	.885
77	Markovits & Thompson (2008) Study 2 [factors inverted]	Argument evaluation tasks	Percentages of acceptance / probability ratings – 7-year-olds	Percentages of acceptance / probability ratings – 9-year-olds	1) Categorical / probabilistic instructions 2) High / low probability 3) Deductive first / probabilistic first	–	**.013**	.309	.594

78	Obersteiner, Bernhard, & Reiss (2015) Study 2	Contingency table problems	Proportion correct – 8-year-olds	Proportion correct – 10-year-olds	1) *a*-versus-*c* / additive / multiplicative	1.000	–	–	–

Note. Significant *p*-values ($p < .05$) are in bold font. CWRAT = Chinese Word Remote Associate Test; MP = Modus Ponens; AC = Affirmation of the Consequent.

Chapter 4: Discussion

4.1 Overview

The aim of the current work was to test the competing single-process and dual-process accounts of reasoning in a developmental context. This study built a database consisting of 78 datasets from 10 published developmental studies compiled using a systematic search strategy to which dual-process accounts have been applied. State-trace analysis was used to re-analyse the data to determine whether they best support the single- or dual-process accounts and to identify experiment factors or designs that can lead to promising evidence for dual-process accounts of reasoning. A key finding was that the majority of developmental data were consistent with a single underlying latent variable, presenting limited support for dual-process accounts of reasoning[6]. Notably, compared to chronological age, experiment factors concerning reasoning task characteristics or additional experimental manipulations were found to be more potent factors as dependent variables in state-trace analysis leading to evidence in favour of dual processes in reasoning development.

4.2 Current findings

4.2.1 Evidence in favour of dual-process accounts of reasoning

The current study found some evidence to reject the simplest single-process account with one underlying latent variable, thereby providing some evidence for dual-process accounts of reasoning. According to the STA results nearly a quarter of the datasets (24%) showed statistical significance, suggesting more than one latent variable – note that these results were obtained when more realistic *SD*s of .30 were assumed for datasets with unknown *SD*s. The percentage of the significant state-trace results was as high as 37% when *SD*s of .10 were assumed, suggesting that more evidence of multiple latent variables might

[6] Unless otherwise indicated, the discussion in this chapter was based on STA results of datasets with known *SD*s and unknown *SD*s when *SD*s of .30 were assumed.

exist in the developmental reasoning data. Although possible, such results were unlikely to approximate reality considering that the mean reported *SD*s in the database was .26. Moreover, assuming *SD*s of .30 for datasets with unknown *SD*s, none of the datasets (including those with known *SD*s) in the Stephens et al. (2019) database of adult reasoning studies showed evidence of multiple latent variables. In comparison, it appeared that evidence favouring multiple psychological processes was more pronounced in the developmental reasoning data than the adult reasoning data. Nevertheless, the evidence against one latent variable from this study should be interpreted with caution, considering that such evidence was mainly derived from a single study – Lin and Shih (2016).

Interestingly, the Lin and Shih (2016) data contributed to 15 out of the 19 significant state-trace results of this study. The authors have evaluated a "dual-process account of creativity" theory, which hypothesised that open-ended divergent thinking and close-ended convergent thinking have different involvement of Type 1 and Type 2 processing. To some extent, the positive results provided favourable evidence for both the "dual-process account of creativity" theory (see Lin & Lien, 2013) and the more general dual-process account of reasoning (e.g., Evans & Stanovich, 2013) upon which the creativity theory was developed. However, the Lin and Shih data might have provided misleading evidence in support of the dual-process accounts. The design of the Lin and Shih experiment was distinct from typical reasoning experiments (cf. e.g., Markovits et al., 2019), in which common stimuli and other extraneous design features are held constant across two measures thought to differentially capture Type 1 versus Type 2 processing. Instead of controlling experimental factors in this manner, open- and closed-ended thinking was tested using completely different tasks with different instructions, stimuli, and scoring methods. Open-ended thinking was assessed using a divergent thinking test, whereas closed-ended thinking was evaluated using an insight problem, a word association task, and a syllogism reasoning task. The lack of proper experimental control left open the possibility that the two types of thinking processes being tested were not isolated. This allowed potential confounds (e.g., different task demands or stimuli) to potentially affect the experimental results and made it difficult to distinguish the factor that has driven the effects. Therefore, the evidence of multiple latent variables from Lin and Shih could not convincingly rule out a single-process account.

The other 4 significant state-trace results were from datasets 32, 61, 64 and 70, providing potential evidence for dual-process models. Caution is needed in interpreting some of the evidence, given the inconsistent results across datasets from the same experiment. Out of the 16 Klaczynski (2000) datasets, only dataset 32 returned a significant result, which might simply be a Type 1 error. Similarly, when age defined the STA dependent variables, only dataset 64 from Markovits et al. (2019) showed evidence of multiple latent variables. Nonetheless, there was some good evidence in support of two-process accounts – or at least evidence against the simplest single-process account with one latent variable. As a comparison of fast and slow thinking, the two-dimensional state-trace of dataset 61 lent some support for the view of Markovits et al. in the existence of two simultaneously developing reasoning processes. Additionally, comparing less resource-demanding and resource-intensive responses, the result of dataset 70 potentially supported the argument of Markovits and Thompson (2008) that non-use versus deliberate use of counterexamples in conditional reasoning are good indicators of the involvement of two distinct reasoning processes.

4.2.2 Evidence in favour of single-process accounts of reasoning

Crucially, much of the developmental data were shown to be consistent with a single underlying latent variable, as one-dimensionality could not be rejected for the majority of the datasets (76% overall; and 93% if excluding the Lin and Shih (2016) datasets). As such, a simple single-process account with one latent variable should be preferred to account for most of these reasoning data on the grounds of parsimony (see Vandekerckhove, Matzke, &

Wagenmakers, 2015). Although the developmental data to varying degrees have been explained in dual-process terms in the original studies, the current finding suggests that dual-process explanations were largely unnecessary. This finding was more in line with the previous demonstrations that dissociation evidence for qualitatively different processes from adult reasoning data did not stand up against the strict criteria of STA (Stephens et al., 2019).

It is necessary to note that half of the studies included in the database have adopted two important experimental paradigms in the study of reasoning – the belief bias and base-rate neglect paradigms. These two paradigms have been used to test the predictions drawn from dual-process accounts of reasoning, and relevant studies have provided substantial evidence supporting the dual-process accounts (Handley and Trippas, 2015). Surprisingly, the data from the two belief bias studies in the database were largely found to be consistent with a single latent variable. Subjective strength of conclusions in the argument evaluation tasks may be the latent variable that typically compels the changes in the performance across experiment conditions (Stephens et al., 2019).

More surprisingly, none of the three base-rate neglect studies in the database have detected evidence of multiple latent variables. In terms of reasoning with base rates, dual-process theories postulate that Type 1 processing cues the stereotype-based responses, whereas Type 2 processing is required to accurately provide normative base-rate responses (e.g., Evans, 2010). The De Neys and Vanderputte (2011) study in the database was a typical example of this paradigm. Given the significant interaction effects reported by the authors, there was evidence from the original statistical analyses for both intuitive and analytic operations on the reasoning processes. A plausible explanation of why the De Neys and Vanderputte datasets returned null STA results was that the experimental evidence for two separate processes might have relied upon the linearity assumption, thereby failing to hold up under a more realistic monotonic assumption (Stephens et al., 2019). In this regard, the data did not rule out an alternative single-process account under which a latent variable drives the changes in the reasoning performance across different experiment conditions. Such a latent variable could be the relative strength of response options in the base-rate tasks.

Lastly, the current study has included a wider range of reasoning tasks in the database, such as law of large number problems (Klaczynski, 2000) or contingency table problems (Obersteiner et al., 2015). As with the data from studies using the aforementioned paradigms, the data from studies using other thinking tasks have also shown little evidence against a latent variable. Weighing up both the significant and null STA results, this study found that across a range of tasks, single-process models show much promise in accounting for reasoning data that have been explained by more complex dual-process accounts.

4.2.3 Experiment factors leading to evidence for dual processes in reasoning

According to the results of the state-trace analyses, factors concerning experimental manipulations were more potent than age groups as dependent variables in STA at identifying evidence favouring dual processes in reasoning development. Consistent with dual-process theories (e.g., Evans, 2008; Evans & Stanovich, 2013), this finding presented some support that stronger forms of evidence for dual-process accounts come from direct experimental manipulations designed to dissociate Type 1 and Type 2 processing. Additional experimental manipulations (e.g., un-speeded versus speeded conditions) appeared to probe better for evidence favouring two reasoning processes than task characteristics (e.g., familiar versus unfamiliar stereotype cues in base-rate tasks). If low *SD*s of .10 were assumed, evidence of multiple latent variables was found in only half (51%) of the datasets that used task characteristics to define the dependent variables in STA, but in *all* of the datasets that used additional experimental manipulations to define the dependent variables. These additional manipulations included using deductive reasoning instructions thought to increase Type 2 processing effort (Markovits & Thompson, 2008) or imposing time constraints thought to inhibit Type 2 processing (Markovits et al., 2019).

However, this comparison between the two subtypes of dependent variables might not be informative – the problem was that few datasets included factors concerning additional experimental manipulations that could be used as dependent variables in state-trace analysis. There were only 4 such datasets in the database, whereas there were 40 datasets that used factors concerning task characteristics as the dependent variables. Furthermore, if assuming *SD*s of .30, only one dataset that used additional manipulation as dependent variables has indicated evidence against one underlying latent variable – dataset 61 from Markovits et al. (2019). The positive state-trace result of dataset 61 demonstrated that un-speeded/speeded judgement was an effective experiment factor, leading to promising evidence for dual-process accounts of reasoning (see Figure 6). Also, it provided some support for the idea that reasoners would engage in a qualitatively different kinds of processing when they are allowed sufficient versus little time for reflective thought (Evans & Stanovich, 2013).

Although developmental studies have found age-related performance increases or decreases on different reasoning tasks and discussed findings from a dual-process perspective (e.g., De Neys & Vanderputte, 2011; Felmban & Klaczynski, 2019), age groups were not found to be potent factors as STA dependent variables in probing for evidence favouring dual processes. Unexpectedly, evidence against one latent variable could only be found in 1 out of the 34 datasets where age defined the dependent variables. A plausible explanation for age failing to capture the distinction between separate reasoning processes in the developmental data is that age is an indirect measure of cognitive ability (Kokis, Macpherson, Toplak, West, & Stanovich, 2002). Cognitive ability is shown to highly correlate with Type 2 processing and is considered as a key factor in differentiating the two types of mental processing in dual-process theories (Evans & Stanovich, 2013). Another plausible explanation is that the aggregate data used might distort the picture that could emerge from STA, as they are an average across groups of individuals (Stephens et al., 2019). Given the proposed individual differences in the development of Type 1 and Type 2 processing (see Stanovich et al., 2011), age may play a better role in tapping into the Type 1/2 dichotomy and detecting evidence for distinct processes had the data been analysed at an individual level (i.e., with longitudinal data), or with subgroups based on cognitive ability at each age level.

4.3 Strengths and limitations

The current work can be seen as an example of how the field of reasoning development can benefit from the testing of single-process and dual-process accounts using a more rigorous method – state-trace analysis. Assuming a monotonic relationship between latent and dependent variables, STA is a more stringent inferential tool to use for testing whether the dependent variables are influenced by more than one latent variable (Newell & Dunn, 2008). Importantly, the use of STA could help address the problem of high Type 1 error rates inherent in the use of linear statistical tools such as ANOVA when drawing conclusions about multiple psychological processes – a significant ANOVA interaction is often found when the state-trace result is consistent with a single latent variable, due to the assumption of a linear relationship between latent and dependent variables (Dunn et al., 2014). Therefore, the application of STA on the reasoning data that were previously analysed with linear models allowed the current work to contribute less Type 1 error-prone evidence to theorizing about cognitive mechanism(s) underlying reasoning.

The application of STA also revealed limitations of this study. Care must be taken in interpreting the STA results. When the null hypothesis of a single underlying latent variable is retained for a given dataset, the result does not prove a single-process account or conclusively dismiss the possibility of more than one underlying psychological process (Stephens et al., 2019). Consider, for example, that a one-dimensional state-trace could result from the fact that two dependent variables across conditions were influenced by two latent variables in the same way. Indeed, when it comes to cases where the null hypothesis cannot be rejected, STA can only warrant conclusions that there is no evidence in the data that compels multiple mental processes (Stephens et al., 2019). Alternatively, evidence of a two-dimensional state-trace may also be consistent with a complex single-process account in which multiple parameters are proposed (e.g., argument strength and response bias parameters; see Stephens et al., 2019). However, STA is a useful initial test of the simplest single-process account – and the current results indicate that often this account is sufficient.

In addition, the STA CMR test is within the confines of a frequentist framework. For one thing, this study could not use the state-trace results to quantify evidence favouring the null hypothesis of one underlying latent variable (Stephens et al., 2019). For another, with the alpha level set at .05, it was possible that some statistically significant results revealed in this study were Type 1 errors, especially when a total of 110 CMR tests have been run and multiple tests have been applied to the same experiment data.

Last but not least, there were limitations inherent in the data analysed in this study, highlighting the caveats on the STA results. First, the current study was not an exhaustive meta-analysis of all developmental reasoning studies. There were likely additional empirical studies relevant to testing the models that have not been included in the database. However, by focusing on studies that applied dual-process theories, the database was a stronger testbed for detecting evidence against a single latent variable. Second, STA was applied on aggregate data (i.e., data averaged over participants). In this regard, this study could only make claims about rival single- and dual-process accounts at an aggregate level rather than at an individual level. Third, the results of this study were to some extent reliant upon assumed *SD*s. The potentially significant results from Ameel et al. (2007) and Markovits and Thompson (2008) thus warrant future replication studies, conducting similar experiments that are designed for

STA. Also, the use of known or assumed summary *SD*s meant that the exact variability of the relevant data was unable to be captured. The observed data for each dependent variable in each condition were assumed to be normally distributed; thus, the STA CMR results have relied on the use of the parametric bootstrapping (rather than non-parametric bootstrapping, which can be applied when the raw data are available). Last, some experiments in the developmental database might have been under-powered against the criteria of STA, especially for those with fewer experiment conditions (Stephens et al., 2019). For instance, the Obersteiner et al. (2015) experiment has only met the minimal design for STA; its dataset consisted of three conditions and returned a non-significant CMR test result. Given that the chances of detecting two-dimensionality would increase in datasets with more conditions (see Prince, Brown, & Heathcote, 2012), it was not surprising that datasets with a small number of experiment conditions failed to detect evidence of multiple latent variables.

4.4 Implications and future directions

Overall, the current work presents evidence that challenges the status of dual-process theories as the dominant theoretical framework in the study of reasoning. Extending the STA work on adult reasoning data (Stephens et al., 2019), this study found limited evidence favouring dual-process accounts of reasoning from the developmental data. Instead, the state-trace results indicate that single-process accounts are promising in explaining most of the developmental data in the database. This finding contradicts the interpretations of the data in the original studies given, where most were explained in dual-process terms. Also, in screening for eligible studies for STA, it was found that dual-process theories have often been assumed to be a valid theoretical account, without directly considering or comparing predictions from single-process accounts. Therefore, findings of the current work suggest that dual-process accounts might have been bolstered by false positive evidence, while single-process accounts might have been prematurely discounted. In a broader context where psychology is facing a replication crisis (Pashler & Wagenmakers, 2012), this study highlights the need of also using more rigorous statistical tools in the analyses of data. This also has implications for other domains of cognitive science as far as drawing inferences about latent processes from behavioural data is concerned.

According to Evans and Stanovich (2013), strong and converging evidence for dual-process accounts of reasoning comes from experimental manipulations designed to affect one type of processing but not the other. Compared to the amount of empirical work done with adult populations (also see Stephens et al., 2019), reasoning experiments and especially those with additional experimental manipulations to most rigorously test dual-process models were found to be lacking with children or adolescents. Consider that the database search returned a relatively small number of studies (N = 140 after duplicates removal) and only 2 out of the 10 eligible studies have included the experimental manipulations highlighted by Evans and Stanovich. This further indicates that many reasoning experiments were not designed to test the dual-process accounts. To address this research gap, future developmental reasoning studies are recommended to manipulate key factors thought to distinguish Type 1 and Type 2 processing, to properly test the dual-process models. These factors can be low/high working memory load or cognitive ability, un-speeded/speeded judgements, or intuitive/deductive instructions (Stephens et al., 2019). Additionally, researchers could add testing formal models of reasoning to their research questions, evaluating quantitative instantiations of competing theories. Furthermore, the current work reinforces the need for researchers to be more cautious about the underlying assumptions of their statistical tests when drawing inferences. Instead of relying on task dissociations or interaction effects to infer latent process(es), researchers are recommended to design reasoning experiments for STA instead of ANOVA or other linear models; to provide more compelling evidence for dual-process accounts, demonstrating a two-dimensional state-trace should be the target data pattern, rather than a standard interaction effect.

Lastly, given the need to further explore any two-dimensional state-traces revealed in state-trace analysis, signed difference analysis (SDA) is suggested for use in future research to distinguish competing formal, multi-parameter single-process and dual-process models (see Dunn & Anderson, 2018; Dunn & James, 2003). While STA can help determine whether one or more latent variables underlie manifest behavioural data, SDA can help identify the particular latent variables mediating the effects of experimental factors on the behavioural data. In short, these future research suggestions highlight the need of more rigorous evidence for advancing theoretical accounts of reasoning.

4.5 Conclusion

The current study has re-appraised developmental reasoning data from empirical studies to which dual-process accounts have been applied. According to state-trace analysis, only limited evidence was identified for dual-process accounts of reasoning. Much of the data are more consistent with a single-process account with one underlying latent variable. Findings from this study highlight the need to further test the competing single- and dual-process models with better experimental design (e.g., using experimental manipulations that increase the chances of dissociating Type 1 and Type 2 processing) as well as more stringent statistical tools (e.g., STA or SDA). Ultimately, the aim is to provide a stronger evidence base for accounts of the cognitive mechanisms underlying reasoning and its development.

References

*Ameel, E., Verschueren, N., & Schaeken, W. (2007). The relevance of selecting what's relevant: A dual process approach to transitive reasoning with spatial relations. *Thinking & Reasoning, 13*(2), 164–187. doi:10.1080/13546780600780671

Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology, 19,* 137–181. doi:10.1016/0022-2496(79)90016-6

Barrouillet, P. (2011). Dual-process theories and cognitive development: Advances and challenges. *Developmental Review, 31*, 79–85. doi:10.1016/j.dr.2011.07.002

Chiesi, F., Gronchi, G., & Primi, C. (2008). Age-Trend–Related Differences in Tasks Involving Conjunctive Probabilistic Reasoning. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 62*(3), 188–191. doi:10.1037/1196-1961.62.3.188

Chiesi, F., Primi, C., & Morsanyi, K. (2011). Developmental changes in probabilistic reasoning: The role of cognitive capacity, instructions, thinking styles, and relevant knowledge. *Thinking & Reasoning*, *17*(3), 315–350. doi:10.1080/13546783.2011.598401

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17*, 428–433. doi: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*, 28–38. doi:10.1177/1745691611429354.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*(3), 1248–1299. doi:10.1016/j.cognition.2007.06.002

*De Neys, W., & Vanderputte, K. (2011). When less is not always more: Stereotype knowledge and reasoning development. *Developmental Psychology, 47*(2), 432–441. doi:http://dx.doi.org/10.1037/a0021313

Dunn, J. C. (2008). The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review, 115,* 426–446. doi:10.1037/0033-295X.115.2.426

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

Dunn, J. C., & James, R. N. (2003). Signed difference analysis: Theory and application. *Journal of Mathematical Psychology, 47(4*), 389–416. doi:10.1016/S0022-2496(03)00049-X

Dunn, J. C., & Kalish, M. L. (2018). *State-trace analysis*. Springer.

Dunn, J. C., Kalish, M. L., & Newell, B. R. (2014). State-trace analysis can be an appropriate tool for assessing the number of cognitive systems: A reply to Ashby (2014). *Psychonomic Bulletin & Review, 21*, 947–954. doi:10.3758/s13423-014-0637-y

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

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

Evans, J. St. B. T. (2010). *Thinking twice: Two minds in one brain*. Oxford: Oxford University Press.

Evans, J. St. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review, 31*(2-3), 86–102. doi:10.1016/j.dr.2011.07.007

Evans, J. St. B. T., Barston, J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition, 11*(3), 295–306. doi:10.3758/BF03196976

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

Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. (2010). The influence of cognitive ability and instructional set on causal conditional inference. *The Quarterly Journal of Experimental Psychology, 63*, 892–909. doi:10.1080/17470210903111821

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*, 223–241. doi:10.1177/1745691612460685

*Felmban, W. S., & Klaczynski, P. A. (2019). Adolescents' base rate judgments, metastrategic understanding, and stereotype endorsement. *Journal of Experimental Child Psychology, 178*, 60–85. doi:10.1016/j.jecp.2018.09.014

*Hagá, S., Garcia-Marques, L., & Olson, K. R. (2014). Too young to correct: A developmental test of the three-stage model of social inference. *Journal of Personality and Social Psychology, 107*(6), 994–1012. doi:10.1037/pspa0000012

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. doi:10.1016/bs.plm.2014.09.002

Hayes, B., Wei, P., Dunn, J., & Stephens, R. (2020). Why Is Logic So Likeable? A Single-Process Account of Argument Evaluation With Logic and Liking Judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(4), 699–719. doi:10.1037/xlm0000753

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York, NY: Basic Books.

Jacobs, J., & Potenza, M. (1991). The Use of Judgement Heuristics to Make Social and Object Decisions: A Developmental Perspective. *Child Development, 62*(1), 166–178. doi:10.1111/j.1467-8624.1991.tb01522.x

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus & Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge: Cambridge University Press.

*Kail, R. V. (2013). Influences of credibility of testimony and strength of statistical evidence on children's and adolescents' reasoning. *Journal of Experimental Child Psychology, 116*(3), 747–754. doi:10.1016/j.jecp.2013.04.004

Kalish, M. L., Dunn, J. C., Burdakov, O. P., & Sysoev, O. (2016). A statistical test of the equality of latent orders. *Journal of Mathematical Psychology, 70,* 1–11. doi:10.1016/j.jmp.2015.10.004

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science, 4*, 533–550. doi:10.1111/j.1745-6924.2009.01164.x

*Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development, 71*(5), 1347–1366. doi:10.1111/1467-8624.00232

Klaczynski, P. A. (2009). Cognitive and social cognitive development: Dual-process research and theory. In J. St. B. T. Evans, & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 265–292). Oxford, UK: Oxford University Press.

Kokis, J. V., Macpherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology, 83*(1), 26–52. doi:10.1016/S0022-0965(02)00121-2

Kruglanski, A. (2013). Only One? The Default Interventionist Perspective as a Unimodel—Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science, 8*(3), 242–247. doi:10.1177/1745691613483477

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

Lassiter, D., & Goodman, N. D. (2015). How many kinds of reasoning? Inference, probability, and natural language semantics. *Cognition, 136*, 123–134. doi:10.1016/j.cognition.2014.10.016

Lin, W. L., & Lien, Y. W. (2013). The different role of working memory in open-ended versus closed-ended creative problem solving: A dual-process theory account. *Creative Research Journal*, 25, 85–96. doi:10.1080/10400419.2013.752249

*Lin, W. L., & Shih, Y. L. (2016). The developmental trends of different creative potentials in relation to children's reasoning abilities: From a cognitive theoretical perspective. *Thinking Skills and Creativity, 22*, 36–47. doi:10.1016/j.tsc.2016.08.004

Loftus, G. R. (1978). On interpretation of interactions. *Memory & Cognition, 6*, 312–319. doi:10.3758/BF03197461

Markovits, H. (2017). The development of logical reasoning. In L. J. Ball, & V. A. Thompson (Eds.), *International handbook of thinking and reasoning* (pp. 383–400). New York, NY: Routledge.

*Markovits, H., de Chantal, P. L., Brisson, J., & Gagnon-St-Pierre, É. (2019). The development of fast and slow inferential responding: Evidence for a parallel development of rule-based and belief-based intuitions. *Memory & cognition*, *47*(6), 1188–1200. doi:10.3758/s13421-019-00927-3

*Markovits, H., & Thompson, V. (2008). Different developmental patterns of simple deductive and probabilistic inferential reasoning. *Memory & Cognition, 36*(6), 1066–1078. doi:10.3758/MC.36.6.1066

Morsanyi, K., & Handley, S. J. (2008). How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgment. *Journal of Experimental Child Psychology*, *99*(1), 18–36. doi:10.1016/j.jecp.2007.08.003

Nettelbeck, T., & Burns, N. (2010). Processing speed, working memory and reasoning ability from childhood to old age. *Personality and Individual Differences, 48*(4), 379–384. doi:10.1016/j.paid.2009.10.032

Newell, B. R., & Dunn, J. C. (2008). Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences, 12*, 285–290. doi:10.1016/j.tics.2008.04.009

*Obersteiner, A., Bernhard, M., & Reiss, K. (2015). Primary school children's strategies in solving contingency table problems: the role of intuition and inhibition. *ZDM - Mathematics Education*, *47*(5), 825–836. doi:10.1007/s11858-015-0681-8

Osman, M. (2013). A case study: Dual-process theories of higher cognition—Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science, 8*, 248–252. doi:10.1177/1745691613483475

Overton, W. F. (1990). Competence and procedures: Constraints on the development of logical reasoning. *Reasoning, necessity, and logic: Developmental perspectives*, 1–32.

Ovid Technologies. (2020). Search Tools. Retrieved 14 April 2020, from http://site.ovid.com/site/help/documentation/osp/en/index.htm#CSHID=subjsrch.htm|StartTopic=Content/subjsrch.htm|SkinName=OvidSP_WebHelp_Skin

Pashler, H., & Wagenmakers, E-J. (2012). Editors introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7,* 528–530. doi:10.1177/1745691612465253

Prince, M., Brown, S., & Heathcote, A. (2012). The design and analysis of state-trace experiments. *Psychological Methods, 17,* 78–99. doi:10.1037/a0025809

Reyna, V. F., & Farley, F. (2006). Risk and rationality in adolescent decision making: Implications for theory, practice, and public policy. *Psychological Science in the Public Interest, 7,* 1–44. doi:10.1111/j .1529-1006.2006.00026.x

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

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

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago, IL: University of Chicago Press. doi:10.7208/chicago/9780226771199.001.0001

Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York, NY: Oxford University Press.

Stanovich, K., & Toplak, M. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society, 11*(1), 3–13. doi:10.1007/s11299-011-0093-6

Stanovich, K. E., Toplak, M. E., & West, R. F. (2008). The development of rational thought: A taxonomy of heuristics and biases. *Advances in Child Development and Behavior, 36*, 251–285. doi:10.1016/S0065-2407(08)00006-2

Stanovich, K., & West, R. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*(5), 645–665. doi:10.1017/S0140525X00003435

Stanovich, K., West, R., & Toplak, M. (2011). The complexity of developmental predictions from dual process models. *Developmental Review, 31*(2-3), 103–118. doi:10.1016/j.dr.2011.07.003

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

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*, 104223. doi:10.1016/j.cognition.2020.104223

Stephens, R. G., Matzke, D., & Hayes, B. K. (2019). Disappearing dissociations in Experimental Psychology: Using state-trace analysis to test for multiple processes. *Journal of Mathematical Psychology, 90,* 3–22. doi:10.1016/j.jmp.2018.11.003

Toplak, M., West, R., & Stanovich, K. (2014). Rational Thinking and Cognitive Sophistication: Development, Cognitive Abilities, and Thinking Dispositions. *Developmental Psychology*, *50*(4), 1037–1048. doi:10.1037/a0034910

Vandekerckhove, J., Matzke, D., & Wagenmakers, E-J. (2015). Model comparison and the principle of parsimony. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels

(Eds.) *Oxford handbook of computational and mathematical psychology* (pp. 300–319). Oxford, UK: Oxford University Press.

Venet, M., & Markovits, H. (2001). Understanding Uncertainty with Abstract Conditional Premises. *Merrill-Palmer Quarterly, 47*(1), 74–99. doi:10.1353/mpq.2001.0006

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*, 145–160. doi:10.3758/s13421-011-0158-0

Note: References marked with an asterisk indicate studies included in the reasoning database.