Reasoning and discourse coherence in Autism Spectrum Disorder

Stenning, K.; Van Lambalgen, M.

DOI
10.4324/9781351060912-7

Publication date
2020

Document Version
Submitted manuscript

Published in
Thinking, Reasoning and Decision Making in Autism

REASONING AND DISCOURSE COHERENCE IN AUTISM SPECTRUM DISORDER

KEITH STENNING AND MICHIEL VAN LAMBALGEN

1. Introduction

For diagnostic purposes, Autism Spectrum Disorder is characterised by two clusters of symptoms: problems in social communication and social interaction (including skills requiring 'theory of mind'), and restricted, repetitive patterns of behaviour, interests or activities. It is by now abundantly clear that these symptoms are indicative of fundamental differences with neurotypical cognition and structure of the brain. This chapter shows that people with ASD differ from neurotypicals in some of the logical inferences that they endorse or reject, and argues that this fact is a fundamental feature of autistic cognition, rooted in homologies between language and motor planning, particularly action-goals and the synthesis of action sequences. To substantiate these claims, we first (section 2) develop a view of logical reasoning as consisting of two phases: reasoning to an interpretation, and reasoning from an interpretation.[1] For example, a reasoning task typically involves several premises, but is not explicit about how these premises must be combined. It is usually assumed that the combination rule is the conjunction of classical logic; on the contrary, modern logic has shown there are many options, each relevant to modelling empirical phenomena. One therefore has to 'go beyond the information given' by means of a logic which specifies an interpretation that must include rules for combination of the premises. Once that is done, the reasoning task can be solved by reasoning from the interpretation. The most important takeaway from this discussion is that in a reasoning task, say an instance of modus ponens (MP), the reasoner decides only after settling upon an interpretation of the premises, how MP is going to be applied. This application may be different from the one that suggests itself before interpretation – plenty of examples will be given below.

The focus on interpretation means reasoning tasks belong to the domain of discourse processing. It is known that people with ASD

The authors wish to express their gratitude to Ruth Byrne and Kinga Morsanyi for their vigorous editorial comments.

[1]See the chapter by Geurts et al. (this volume) for an analogous distinction in pragmatics.

have deficits in this domain, and this chapter shows how these deficits shape their reasoning.[2]

Combination of premises is particularly difficult in the case of reasoning with rules which tolerate exceptions that are not counterexamples. Here reasoning to an interpretation is generally non-monotonic: adding a premise may lead to retracting a conclusion. This brings to mind the 'suppression effect' (Byrne, 1989), and the suppression task is the experimental paradigm we will be using in sections 4 and 6. We believe this paradigm shows its true colours only when viewed through the lens of a non-classical logic, *closed world reasoning*. In section 4 we explore (theoretically and experimentally) that part of closed world reasoning that is common to both autists and neurotypicals. Section 6 identifies the point where autistic reasoning diverges from neurotypical reasoning, by showing that tasks which people with ASD find challenging – false belief task and box task – have a logical structure similar to the suppression task. Our diagnosis is that these subjects have a more restricted set of devices for combining premises, due to representational deficits when it comes to exceptions. In section 7 we tie the various threads together and put the results obtained in a broader perspective.

2. Reasoning

To set the stage for this chapter, we begin by outlining the view of reasoning presented in (Stenning & van Lambalgen, 2008). We do so because in our view the experiments reported here, and their logical analysis, make sense only if reasoning is viewed as discourse processing, where by *discourse* we mean a piece of text that can be seen as originating in a communicative interaction, e.g. a conversation. It follows from this definition that a given sentence in a discourse must be interpreted as coherent with what came before, more particularly as adding information. To achieve coherence is not trivial, and generally requires reasoning, what we call 'reasoning to an interpretation.' For example, someone could utter 'Max fell', only to add a bit later 'John pushed him'. The question is: what is the true order of events, as possibly distinct from the sentence ordering? Reasoning to a coherent interpretation of this (mini-)discourse requires as a first step the choice of a language in which to describe the possible relations between the events. Prominent candidates are the language of causes and effects, and the language of goals, plans and action; in this case the former is the most fitting. In view of the question to be answered, the second step is then to collect principles connecting causality and temporal order. In the final step, the hearer has to determine which of (Max fall, John push

[2]The distinction between the two forms of reasoning parallels the distinction in the autism literature between constructing the gist of a narrative and drawing inferences about the narrative making use of the gist (Diehl, Bennetto, & Young, 2006).

–) is cause, and which is effect. After the subsumption of 'push' under cause and 'fall' under effect the interpretation is complete. Reasoning *from* an interpretation then answers the initial question, by means of the principle that causes precede effects:

(1) John pushed Max, and then Max fell.

Although this example is simple, it illustrates a general principle: often the full stop between two utterances is not a logical conjunction, but a connection called 'rhetorical relation' – one of which is causality. Other examples of rhetorical relations are (see (Asher & Lascarides, 2003)): explanation, elaboration, (necessary or sufficient) condition, goal/plan, motivation, intention, inference .... On a discourse approach to reasoning, extracting rhetorical relations between premises using material retrieved from semantic memory is part and parcel of the subject's reasoning. As noted above, we call the type of reasoning issuing in such enriched semantic representations 'reasoning to an interpretation [of the discourse]'. The complementary process, 'reasoning from an interpretation', does not involve memory retrieval, but takes the verbal discourse and its (enriched) interpretation as input. Reasoning from an interpretation may involve inferences *in* the logic determined by the interpretation as well as inferences *about* the logic, e.g. that some proposition is not derivable.

We are now in a position to give a working definition of 'discourse': a *sequence* of sentences, linked by various rhetorical relations detailing how each sentence adds a piece of information to the information made available by its predecessors. This definition is intended to cover examples ranging from narratives to (mathematical or informal) proofs.The emphasis on linking is relevant for our purpose because our main experimental paradigm – Byrne's suppression task (Byrne, 1989) – involves linking the premises going beyond a simple conjunction. For this reason it is of interest to administer the suppression task to subjects with ASD, since these subjects tend to have difficulties with linking the events described by sentences in a discourse. Data from (Tager-Flusberg, 1995) show that children with ASD engaged in narrating the wordless picture book *Frog, Where Are You* (Mayer, 1969) overwhelmingly failed to link events using causal statements. More generally, Loth, Gómez, and Happé (2008) showed that adults with ASD cannot deal very well with the hierarchical organisation of event schemas: '[E]xpectations of what happens in familiar events may be unusually concrete and rigid, which impacts in particular on the understanding of variable aspects of social experiences.' As we will see below, the suppression task is concerned with reasoning about event schemas with many variable aspects.

Reasoning to an interpretation is a non-monotonic process tracking the linguistic input, and may lead to a reversal of the event ordering. Let e denote the event 'fall', f the event 'push'. Suppose 'John pushed

him' is not a complete sentence, but rather the initial part of 'John pushed him into the hole dug expressly for the purpose.', then the resulting event ordering ($e < f$) is congruent to the sentence ordering. In logical terms, this means that the initial inference to $f < e$ is *non-monotonic*: it may have to be retracted when new information arrives. Another interesting perspective is provided by processing. There is experimental evidence for the claim that a situation model for the discourse 'Max fell. John pushed him into the hole dug expressly for the purpose.' is constructed incrementally, yielding the sequence of models $M_1 = \{e\}$, $M_2 = \{f < e\}$ and $M_3 = \{e < f\}$. While M_2 is an extension of M_1, M_3 is a re-computation of M_2; and these two processes have clearly distinct ERP signatures (see (Baggio, van Lambalgen, & Hagoort, 2008) for an analogous case).

The view of reasoning adopted in the psychology of reasoning community is rather different from the one advocated here. In the psychology of reasoning, the object of study is a particular form of discourse known as 'deductive reasoning'. The task is to generate logically valid conclusions ('necessary conclusions') from given premises, and nothing else. The premises are combined conjunctively (not via rhetorical relations), and a separate stage of interpreting premises and task is deemed unnecessary, indeed highly undesirable, because interpretation would add premises. In practice, 'logically valid' is identified with validity in classical logic, which is, in the psychology of reasoning, considered to be the unique normatively justified logic.

Rather than present a formal system for classical logic, we highlight two defining characteristics of classical logic most relevant to conditional reasoning. First of all, classical logic is monotonic – given a valid argument, adding a new premise does not affect the validity of the argument. Monotonicity implies that the classical ('material') implication 'if p then q' must be false if p is true and q is false – i.e. exceptions are counterexamples. Secondly, if we require also *bivalence* – a proposition is either true or false, with nothing in-between – then 'there is no counterexample' is a sufficient condition for the truth of the implication 'if p then q', and we get full classical logic. Why should this logic be the norm? The norm imposes a particular meaning on the conditional, prominent in the intended interpretation of the Wason selection task,[3]but which is not by any stretch of the imagination the common meaning (Declerck & Reed, 2001). Monotonicity is a component of the norm for classical logic with much better credentials, and is connected to the notion of proof. Let Δ be a set of statements proved up to the present moment, and I want to prove 'if p then q' from Δ. One then adds p as an assumption to Δ and tries to derive q. If the logic were non-monotonic, addition of p could lead to a retraction of part of Δ; and one wouldn't know which part of Δ can still be used

[3]See (Stenning & van Lambalgen, 2008, Chapter 3) for discussion.

for deriving q. Therefore a good case can be made for adopting monotonicity in the context of proof; but there are numerous contexts in which reasoning cannot be assumed to be monotonic, as we will see below.

Answers which deviate from the norm are said to be 'irrational', or at the very least, evidence of a cognitive bias. The effect of having one logical norm, supposedly applicable in whichever situation the subject finds herself, is that the actual behaviour, which almost always violates the norm, must be relegated to the non-logical. However, these so-called biases may be the result of subjects reasoning according to a very different norm. This is not like Groucho Marx's 'These are my principles, and if you don't like them, I have others'. Above we showed that monotonicity is a reasonable requirement when dealing with proofs. By contrast, plans may fail because of unknown preconditions, which must be incorporated in the plan once they have been identified, and here there is no room for monotonicity. Proofs and plans are different domains, and logical principles are not analytic, but depend upon the domain.

Therefore, the experimental question cannot be 'Does the subject reasoning correctly according to that logical norm?'. Rather, there are two questions to consider: 'Which norm does the subject adopt? Does the subject's reasoning conform to the norm?'. Empiricism in this stark form doesn't work, of course; the experimenter will have to select a few norms based on her hypotheses about what subjects may be doing. The methodological difference between the 'deductive reasoning' view and ours can be stated as follows. Suppose a subject reasons in conformity with classical logic. On the former view the subject reasons according to the norm and there is nothing to be explained; on our view there is still the question why the subject adopts this particular norm for the domain she finds herself in, and whether it is reasonable for her to do so in that domain. This applies as well to our main experimental paradigm for studying reasoning to and from an interpretation in autism, Byrne's suppression task. We proceed to give a brief introduction to the task.

2.1. **The suppression paradigm.** In a suppression experiment, the subject is presented with two tasks. The first is a simple conditional reasoning problem consisting of a conditional premise, a categorical premise and a putative conclusion. Here are two examples, with typical scores:
A *modus ponens* (MP) argument (which is classically valid):

(2) *If she has an essay to write she will study late in the library.*
She has an essay to write.
and the subjects are asked whether the putative conclusion:
(3) *She will study late in the library.*

is true given that the premises are true. The possible answers are 'yes', 'no', 'maybe'. In this case 96.1% of neurotypical subjects answer 'yes'.[4] An *affirmation of the consequent* (AC) argument (classically invalid)

(4) *If she has an essay to write she will study late in the library.*
She has an essay to write.
and the subjects are asked whether the putative conclusion:
(5) *She has an essay to write.*

is true given that the premises are true. The possible answers are 'yes', 'no', 'maybe'. In this case 67.1% of neurotypical subjects answer 'yes'. In the second task, conditional premises are added to the simple inferences above, and the question is repeated. For MP, the second conditional premise could be of the following type (an 'additional conditional')

(6) *If the library is open, she will study late in the library.*

The percentage of 'yes' answers drops to 51.1%.
In the case of AC, one is interested in 'alternative conditional' as third premise[5]

(7) *If she has a textbook to read, she will study late in the library.*

The percentage of 'yes' answers drops to 9.6%. In Byrne's original formulation, this pattern is said to show *suppression* of MP and AC.

2.1.1. *Is suppression evidence for context sensitivity?* On the 'deductive reasoning' view, a subject who answers 'maybe' in the three-premise MP task, instead of 'yes', is guilty of a sin of omission, whereas a 'maybe' answer in the three premise AC task, indicates correct logical reasoning. In classical logic it makes sense to distinguish between premises and context: since MP is a two-premise inference, the third premise can be considered context, which, by the monotonicity of classical logic, should not affect the endorsement of MP. This line of reasoning underlies a series of experiments by McKenzie, Evans, and Handley (2011) aiming to show that people with ASD are less sensitive to context than neurotypicals. Byrne's suppression task had shown that neurotypical subjects engaged in MP inference endorse MP to a significantly lesser degree in the presence of an additional premise taken as context. If one could show that subjects with ASD endorse MP significantly more often in the presence of an additional premise, then this would be evidence for lack of sensitivity to context in these subjects. This argument is flawed, but it is so in interesting ways. Consider

[4]All data taken from (Pijnacker et al., 2009), data reproduced below as figure 1.
[5]The full experiment treats also MT and DA, as well as AC with alternative second conditional, etc.

first the important, but notoriously underspecified notion of context. A minimum requirement is the possibility to distinguish empirically between intrinsic task demands and extrinsic factors that may influence task performance. In fact, the distinction depends on one's view of human reasoning and its relation to logic. If one views reasoning, say applying MP, as (i) operating on the surface structure of the given verbal argument, and (ii) obeying the norms of classical logic, then context can be defined as verbal material not necessary to draw the MP inference: by (i) this material is syntactically identifiable, and by (ii), in particular the monotonicity of classical logic, this material should not affect the MP inference. If context thus defined has an effect on a subject's rate of endorsement of MP, one says this subject is *context sensitive*; if not one could say the subject is *context blind*.

2.1.2. *Discourse analysis of suppression.* On a discourse view of reasoning, the three premises in (say) the additional condition are first processed aiming at a coherent interpretation ('reasoning to an interpretation'), and once an interpretation is established, the applicability of MP is investigated ('reasoning from an interpretation'). Here there is no principled distinction between premises and context. Instead, the issue is one of *integration* or *coherence*: high rates of endorsement of MP means the additional premise is only loosely integrated with the other premises (typically with the conjunction 'and'); low rates of endorsement indicate tight integration of the additional premise. On our view, response patterns in the suppression task therefore yield information on the 'degree of coherence' of this discourse. Admittedly, we haven't even gestured at a definition of this concept, but McKenzie, Evans, and Handley (2010) provide a clue. We capture the effect of context at the level of event structure, by giving the first conditional an event parameter e (for 'exception'; we used the notation *ab* ('abnormality') 'in (Stenning & van Lambalgen, 2008)), whose value is determined by an *interpretation* $I(e)$: a list of expressions of the form 'if *not-r* then e'. For concrete r (say 'the library is open'), this formula expresses that '*not-r*' describes an exception or disabling condition. It occurs as the antecedent in an implication because we may wish to leave it open whether '*not-r*' is true or not. The role of the negation in '*not-r*' will be clarified shortly.

The underlying logical form of the first conditional in the suppression task is then given by

(8) a. If she has an essay and *not-e*, then she studies late in the library
b. to be evaluated in interpretation $I(e)$

Let p = 'she has an essay', q = 'she studies late in the library', and r = 'the library is open'. The second conditional

(9) If the library is open, she studies late in the library.

adds to the interpretation as follows. This conditional is likely to imply its obverse (see (Fillenbaum, 1978) for the relevant data)

(10) If the library is closed, she won't study late in the the library.

or in concise form

(11) if *not-r*, then *not-q*.

We then have: if *not-r*, then e.[6] The introduction of the parameter e which is not present in the surface structure of the conditional allows us to integrate the two conditionals much more deeply than what can be achieved using logical connectives only. An even stronger form of integration will be presented in section 2.1.4

2.1.3. *Suppression explained as non-monotonicity.* Let's now have a look at the logic of e. A general principle governing exceptions to a rule is:

(12) There can only be a relatively small number of exceptions to a rule, otherwise there's no rule. This leads to the 'closed world assumption for potential exceptions': let 'if $not\text{-}r_i$, then e' (where $1 \leq i \leq n$) be a list of all known potential exceptions; then one may assume these are all the potential exceptions.[7]

The second principle applies in particular to causal anomalies, e.g. exceptions to the causal rule 'if you push the brake pedal, the car will slow down':

(13) The occurrence of an exception of the type 'causal anomaly' is a rare event. This leads to the 'closed world assumption for actual exceptions': if there is no reason to assume *not-r*, assume *not-not-r*, that is, r.

In terms of our running example: it is provable that 'the library is closed' is a potential exception, but this event is not a causal anomaly, but an event which happens according to a schedule that may be unknown to us. Principle (13) therefore does not apply. Instead we argue as follows.

Assume for simplicity that the initial interpretation is empty: the contribution of the second conditional has not yet been incorporated. Then there is no way to make e true, which means *not-e* is true, and

[6]Sketch of proof: assume *not-r*, then *not-q*, whence *not-*(p and *not-e*), and since p holds, also e.

[7]This can be expressed formally as follows: e if and only if ($not\text{-}r_1$ or ... or $not\text{-}r_n$).

MP applies. Taking into account the second conditional, and putting r = 'the library is open', the updated interpretation $I(e)$ contains (only) 'if *not-r* then e.' It follows from principle (12) that '*not-r* if and only if e.' Since it is possible that the library is closed, e might be true, i.e. *not-e* might be false, and MP no longer applies.

2.1.4. *Synthesis of a new rule.* The following dialogue (taken from (Stenning & van Lambalgen, 2008, Chapter 7)) shows that taking the obverse of the additional premise really occurs, and that this neurotypical subject uses it to produce a new rule from the two given conditionals

> S: Ok yeah I think it is likely that she stays late in the library tonight, but it depends if the library is open. . . so perhaps I think [pauses]. yeah, in a way I think hmm what does it say to me? I mean the fact that you first say that she has an essay to write then she stays late in the library, but then you add to it if the library stays open she stays late in the library so perhaps she's not actually in the library tonight, because the library's not open. I don't think it's a very good way of putting it.
> E: How would you put it?
> S: I would say, if Marian has an essay to write, and the library stays open late, then she does stay late in the library.

In fact, considering the logical form (8) of the main conditional, reproduced here for convenience

(14) a. If she has an essay and *not-e*, then she studies late in the library
b. to be evaluated in interpretation $I(e)$

and the equivalence '*not-r* if and only if e' derived by closed world reasoning, substitution for e gives

(15) If she has an essay and the library is open, then she studies late in the library,

We will see that this is a form of integration which is generally beyond the cognitive capacities of subjects with ASD, but we first direct our attention to types of integration which are accessible to these subjects.

3. Fallacies? the true meaning of AC and DA

The inference 'affirmation of the consequent' (AC): 'if p then q; q; therefore p,' is classically non-valid because given the semantics for the material implication, the premises of the AC inference are true if q is true but p is false. However, it is doubtful whether subjects who endorse AC have this interpretation in mind; the voluminous linguistic literature on conditionals suggests a wealth of other meanings. However, this makes the inference patterns AC and DA all the more

interesting, since they express particular cognitive 'stances'. We will discuss two examples of such stances: causal stance and teleological stance.

3.1. **The causal stance.** Defined by Alvarez and Booth (2015) to be the 'drive for causal understanding', its guiding principle is: 'every event has a cause'. A consequence of this principle is a form of AC (and likewise DA) for causal conditionals. Suppose the subject interprets 'if p then q' (not as a material implication but) as a causal conditional, where p is the cause and q its effect. Assume q is true, one can conclude p provided
(i) every effect must have a cause
(ii) 'if p then q' is the only known causal relation with q as effect.
For then one may argue: q is true, so it must have a cause (by (i)); but 'if p then q' is the only known causal conditional with effect q, therefore p must be the cause. If (ii) fails, with 'if r then q' another known causal relation with q as effect, then the conclusion p is no longer warranted. This argument is an instance of the *closed world assumption*, which says that if a statement is not currently known to be true, one may assume it to be false. The statement at issue here is: 'there is a causal relation 'if r then q' with r different from p.' Reasoning with the closed world assumption is *non-monotonic*: addition of a premise may invalidate an earlier conclusion. According to Congiu, Schlottmann, and Ray (2010) people with ASD are generally not impaired and often superior in understanding (mechanical) causality. One would therefore expect them to do well on non-monotonic reasoning with causal conditionals.

3.2. **The teleological stance.** Csibra and Gergely (2007) proposed that children (with or without ASD) reason according to the *principle of rationality*: they expect agents actions to be directed to a goal and to be the most functional way to achieve the goal within the constraints of the situation. In short, every action is interpreted as directed toward satisfying an associated goal: this is the 'teleological stance'. One easily sees that, given an observed action a, and a conditional 'if *Goal* then a,' inferring that *Goal* is indeed the goal of executing a, is an instance of affirmation of the consequent, as explained above for causality. If there is only one implication 'if *Goal* then a' for given action a, then the teleological stance together with a closed world assumption for goals implies that G is the goal of a.

If there happens to be another conditional 'if H then a' (H a goal) which is activated in working memory together with 'if *Goal* then a, then the teleological stance and closed world reasoning imply that a is directed to at least one of $Goal, H$, but that is all the information that can be extracted from the data.

In our running example, the main conditional is of the form '*if* `goal` *then* `action`, and the materials for the suppression task contain many

more examples. Children with ASD appear to have intact goal attribution (Vivanti et al., 2011), (Csibra & Gergely, 2007). These considerations lead to the prediction that people with ASD and neurotypical controls do not differ on endorsement and suppression of AC when reasoning with goal-action conditionals.

The upshot of the preceding considerations is that for non-monotonic AC/DA reasoning with the types of conditionals discussed is a valid argument pattern, and that *because people with ASD generally have the relevant cognitive abilities or stances*, no significant difference between subjects with ASD and neurotypical controls is to be expected.

4. Defeasible inferences which do not differentiate between ASD and neurotypical subjects: AC and DA

The results reported here can be found in (Pijnacker et al., 2009).[8]

The first condition presents a subject with the following premises:

(16) a. *If she has an essay to write she will study late in the library.*
b. *She studies late in the library.*

67.1% of (neurotypical) subjects draw the conclusion: 'She has an essay to write'.

The second condition adds the conditional ('alternative') premise

(17) *If she has a textbook to read, she will study late in the library.*

In this case, only 9.6% conclude 'She has an essay to write'.

Given the cognitive rigidity that is held to be characteristic of people with ASD, it comes as a surprise that in the ASD group the percentages are 45% and 12.2%.[9] The same pattern holds for DA. When presented with

(18) a. *If she has an essay to write she will study late in the library.*
b. *She has no essay to write.*

69.1% of neurotypicals conclude 'She will not study late in the library'; and this percentage drops to 10.4% if the premise

[8]Participants in this study included 28 high-functioning adults with autism (autistic disorder (HFA), n=11 and Asperger syndrome, n=17) and 28 matched controls, aged 18 to 40 years. Details of the experimental procedure can be found in (Pijnacker et al., 2009). While that paper focussed on reasoning with exceptions, we (KS, MvL) now believe its implications are more general.

[9]There is obviously a difference in the rate of endorsement of (simple) AC between controls and autists: 67.1% versus 45%. The scores for people with ASD do not differ significantly from those of controls after conditionalising on rate of endorsement.

	ASD			Control		
% responses	yes	no	maybe	yes	no	maybe
MP	89.6	0.0	10.4	96.1	2.5	1.4
MP add	71.0	1.1	28.0	51.1	0.7	48.2
MP alt	92.9	0.4	6.8	97.5	0.7	1.8
MT	1.4	79.6	19.0	2.5	92.8	4.7
MT add	0.7	62.1	37.1	0.7	45.0	54.3
MT alt	0.4	90.3	9.3	1.1	95.0	3.9
AC	45.0	1.1	53.9	67.1	2.1	30.7
AC add	28.1	1.1	70.9	35.7	0.0	64.3
AC alt	12.2	2.2	85.7	9.6	0.0	90.4
DA	1.1	48.0	50.9	0.4	69.1	30.6
DA add	2.9	28.9	68.2	2.5	33.6	63.9
DA alt	3.2	15.7	81.1	1.1	10.4	88.5

FIGURE 1. Proportion of responses for simple task and suppression task; 'add' = additional premise, 'alt' = alternative premise

(19) *If she has a textbook to read, she will study late in the library.*

is added. In the ASD group the percentages are 48% and 15.7%. For both AC and DA no significant differences were found between groups in the conditional probability of suppression of AC/DA inferences given the endorsed inferences in the simple task ($p > 0.1$). The conclusion seems to be that people with ASD are as flexible as neurotypicals when it comes to dealing with alternative rules; expressed more formally: that the two experimental groups synthesise the same conditional from the two given conditionals. This representation ties the two conditionals together as in the underlined expression in the argument (20):

(20) If p then q, *but* if r, then *also* q. Therefore p does not follow from q.

Here, *but* indicates that the default inference 'q; if p then q; therefore p' is overridden.

4.1. **Da capo, with additional conditional.** Interesting things happen if the alternative conditional is replaced by an additional conditional:

(21) *When the library is open, she studies late in the library.*

For neurotypicals the rate of endorsement drops from 67.1% to 35.7%, and for subjects with ASD there was a drop from 45% to 28%. After conditionalisation that difference is not significant. But what is

striking is that the drops are much less than in the case of the alternative conditional. Hence both groups have sizable subgroups that have distinguishable mental representations for the alternative and for the additional case. However, as we shall see in section 6, these representations must be different for autists and for neurotypicals.

5. False belief tasks, executive function tasks and their relation to the suppression task

To derive predictions for autists' behaviour in the MP/MT conditions of the suppression task, we show that the logical form of the suppression task is present in the false belief task as well as in executive function tasks, which feature a prepotent response that must be inhibited ('suppressed').

5.1. **False belief and suppression.** In the original version of the false belief task (Wimmer & Perner, 1983), Maxi puts his chocolate in the kitchen cupboard and leaves the room to play. While he is away, his mother moves the chocolate from the cupboard to the drawer. Maxi returns. Where will he look for his chocolate, in the drawer or in the cupboard? Neurotypical children younger than 4yrs answer 'in the drawer', older children answer 'in the cupboard'. Children with ASD of verbal mental age 4–5 fail to give the correct answer. We focus here on an executive function deficit aspects of the results of the ASD group. A task tapping into executive function has two components:
(i) the participant has to suppress a prepotent but incorrect response
(ii) he or she must retain action-relevant information in working memory while doing so.

Thus, in the false belief task it is necessary to suppress one's own true belief whilst simultaneously holding in mind the requirement to answer a question about what the protagonist believes or will do. Before the child gives a verbal reply to the question, it must formulate a verbal rule capturing the gist of the situation (Russell, Hala, & Hill, 2003). Interestingly, properties (i) and (ii) can be cast in logical form such that the result is a system of rules as in the suppression task. As it turns out, extracting a single rule is much more difficult for subjects with ASD than for neurotypicals.

Reasoning to an interpretation of the task yields the following. Conditional (1) gives the prepotent response (property (i)):
(1) If the chocolate is at location l, then report l
Conditional (2) represents task understanding (property (ii)):
(2) If Maxi believes the chocolate is at m, then report m
Let t be the true location of the chocolate. To avoid a clash between rules (1) and (2) for $t = l = m$, a neurotypical child's reasoning to an interpretation allows (1) to be inhibited by insertion of an exception parameter

(1e) If the chocolate is at location t and *not-e*, then report t.
(2) is replaced by its obverse (2o) If Maxi does not believe the chocolate is at t, do not report t.
(1e) and (2o) together imply (3) If Maxi does not believe the chocolate is at t, then e.
Since it can be computed (using closed world reasoning and a causal theory relating perception and belief) that in fact Maxi believes the chocolate is at some $f \neq t$,and hence not at t, the exception e holds, and hence the prepotent response (1e) is not triggered, i.e. suppressed (reasoning from the interpretation). Rule (2) can now be applied with $f = m$, and it follows the subject must report location f (idem).

If the subject cannot integrate (1) and (2) by inserting an exception parameter in (1), then (1) and (2) are in competition. A response based on the application of MP to (1) and the subject's perceptual knowledge, is likely to be faster than a response based on (2), which requires computation of Maxi's belief, using the causal relation between perception and belief. Thus without integration of conditionals the child will give the prepotent verbal answer based on (1). But it is very well possible that (2) – when all computations have been done – will eventually lead to a correct non-verbal 'looking response' (Onishi & Baillargeon, 2005).

5.2. **Inhibition and suppression.** A reconceptualisation of the false belief task as including a suppression task for MP suggests that people with ASD will suppress MP much less than neurotypical controls. This conjecture receives further support from an experiment designed to test the executive dysfunction theory of autism (Hughes & Russell, 1993). This experiment illustrates the difficulties that people with ASD experience in synthesizing a new rule necessary to solve a goal-directed task. The subject's task was to obtain a marble that rested on a platform inside a box (see figure 2; the following description is taken from (Hughes & Russell, 1993)). A direct reach toward the marble broke an infrared light beam, which caused a photoelectric cell inside the box to make the marble drop out of sight through a trap door in the platform. There were two ways in which the marble could be obtained. The subject could turn a knob on the right-hand side of the box to cause a paddle behind the marble to flip it down a chute and out of the box toward the subject: the 'knob route.' Alternatively, the subject could use a switch on the other (left) side of the box to extinguish the light beam and thus enable a reach: the 'switch route.'

The session began when the experimenter invited the subject to take the marble out of the box. After the subject had failed to do this twice because his or her reach triggered the trap-door mechanism, the experimenter demonstrated that a direct route would invariably cause the marble to drop away. The experimenter then demonstrated the

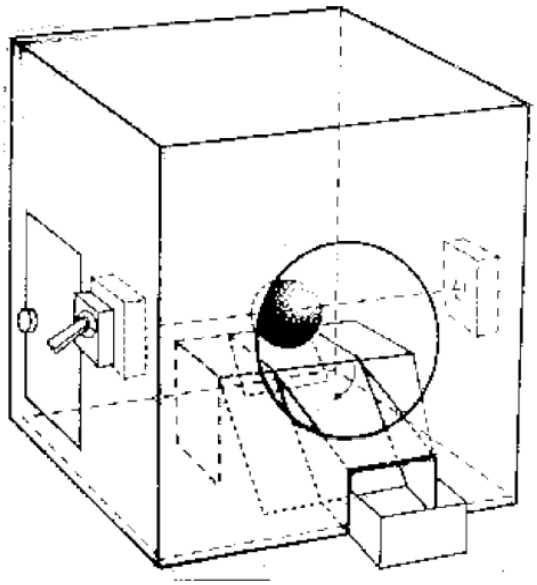

FIGURE 2. Box task

switch route and told the subject that it was now essential to flip the switch before reaching inside the box. She also demonstrated again the consequences of a direct reach. The instructions for the switch route were as follows: 'The way to win the marble is by using this switch here; then you can reach in and get the marble. But you have to make sure that the switch is down first.' Neurotypical children age 4 and older are able to incorporate the switch-instruction into their motor plan; autistic children persevere with the direct reach.

To transform the box task into a suppression task, we formulate two conditionals; we postulate that the neurotypical child of 4yrs and older inserts a parameter e in rule (a) (the prepotent response) to make it inhibitable:
(a) If you reach through the opening and *not-e*, you can retrieve the marble.[10]
(b) If the switch is down, you can retrieve the marble.
We then argue as before. Assume you reach through the opening. (b) is replaced by its obverse, which yields 'if the switch is up, you can't retrieve the marble'. Hence if the switch is up, then e; this implication is put in the previously empty interpretation set $I(e)$. To extract a rule which represents the 'gist' of rules (a) and (b), we have to go one step further and apply the *closed world assumption for exceptions*: since there is no mention of another implication of the form 'if ..., then e' we conclude that 'the switch is up' is the only exception,i.e. 'the switch is up if and only if e', so that we arrive at the gist by substitution in (a)

[10]To be interpreted as conditional (8) in section 2.1.

(c) If you reach through the opening and the switch is down, you can retrieve the marble.

The perserverative behaviour of autists might be viewed as a failure to reason toward a parametrised representation of conditional (a).[11] Given that essential parts of the box task and the false belief task turn out to have the structure of the suppression task for MP, we may predict that subjects with ASD will suppress significantly less than neurotypicals. The next section presents the data,[12] and an interpretation of these data.

6. Suppression task: modus ponens in neurotypicals and subjects with ASD

The first task presents subjects with the premises:

(22) *If she has an essay to write she will study late in the library.*
She has an essay to write.
and are asked whether the putative conclusion:

(23) *She will study late in the library.*

is true given that the premises are true. The possible answers are 'yes', 'no', 'maybe'. In this case 96.1% of neurotypical subjects answer 'yes'. The second task adds the *additional* conditional premise

(24) *If the library is open, she will study late in the library.*

In this case, only 51.1% of neurotypical subjects conclude 'She will study late in the library'.

Here, however, subjects with ASD show a strikingly different pattern: 89.6% in the first condition, 71.0% in the second condition; fewer subjects suppress. Taking MP and MT together the difference between subjects with ASD and neurotypicals is significant ($p < 0.028$). Hence we at last have an experimental condition which separates ASDs' handling of exceptions and that of neurotypicals. But what are the mental representations adopted by the two groups, and in what sense are exceptions involved?

[11]There is a certain similarity between our attempt to identify suppression reasoning in executive and false belief tasks, and the 'Cognitive Complexity and Control' (CCC) theory of Frye, Zelazo, and Burack (1998), which operates with incompatible conditionals (same antecedent, different consequent) that are unified into a higher-order ('if... then (if... then)') conditional by inserting a perspective parameter in the left ... position. More precisely, supposing we have two perspectives p_1, p_2 as in the false belief task, the higher order structure is '`IF` p_1 `THEN` if a_1 then c_1 `ELSE` if a_2 then c_2'. This however is not the unification of conditionals called for in the suppression task.

[12]Percentages from figure 1.

6.1. **Neurotypical group.** We hypothesised that this group adopts the enriched representation (8). Let the basic conditional be represented by 'if p and *not-e* then q' and the additional conditional by 'if r then q'. We have seen that the initially empty interpretation $I(e)$, after processing the additional conditional will contain 'if *not-r* then e'. Barring information that r is the case, these data are sufficient to block ('suppress') MP. But as we have seen in section 2.1.4, neurotypical subjects tend to extract a rule from these data: post-experiment interviews showed that neurotypical subjects tend to synthesise these premises as 'if p and r then q.' This synthesis of the premises immediately explains the suppression of MP in the additional premise case: one would need both p and r to conclude q. It also explains the control group's scores for AC and DA with additional conditional premise. Endorsement of AC demands that the two conditional premises are synthesised into one formula; only in this way a unique conclusion is possible. The data for neurotypicals show that AC for a single conditional is endorsed by 67.1%; with an additional conditional this percentage is still 35.7%, much larger than what is seen with an alternative conditional: 9.7%. This means that in some 35% of subjects (those who do not suppress), the putative conclusion p must occur conjunctively, which supports the proposal above that the synthesis of the conditional premises is indeed 'if p and r then q.'

6.2. **ASD group.** Subjects with ASD who draw the modus ponens conclusion cannot have the representation 'if p and r then q.' even though it would be consistent with their scores on AC with additional premise. The following dialogue[13] gives a clue to what might be happening:

> E: But suppose she has an essay, but the library is closed?
> S: Ah , that's also possible.
> E: Well, I'm only asking.
> S: Well, according to these two sentences that's not possible, I think.

This suggests that the first two premises imply that the library must be open. A derivation ('reasoning to an interpretation') could run as follows: consider the additional premise and, arguing toward a contradiction, assume the library is closed, then by DA applied to the additional premise,[14] she doesn't study late in the library. However, MP implies she does study late in the library, therefore the assumption is wrong and the library is open. Hence only the basic conditional remains, and closed world reasoning leads to the endorsement of AC

[13]Taken from an unpublished MA thesis (?, ?) by the psychiatrist Heleen Smid.

[14]What we called above the 'obverse' following Fillenbaum (1978). We have seen that DA is well within an ASD subject's competence.

with additional premise. Note that this representation also implies that MP is endorsed. The same holds for MT which is endorsed by 62.1% with an additional premise, with a base rate of 79.6%. Importantly, the derivation just given does not refer to the exception parameter e. In fact, the experimental results on the suppression task strongly suggest that the representation (8) is not accessible to subjects with ASD. Independent confirmation of this hypothesis comes from McKenzie et al. (2010). Simplifying a bit, this study involved two classes of conditionals, say *Hi* and *Lo*, defined according to whether a conditional had a high or a low number of disabling conditions. The aim was to compare the rate of endorsement of MP applied to conditionals in *Hi* with that of conditionals in *Lo*, where the rate of endorsement was evaluated in neurotypical subjects and subjects with ASD. As expected, in the neurotypical group, the class *Hi* elicited a significantly lower rate of endorsement than the class *Lo*. However, in the ASD group there was no difference at all. This result is all the more surprising since McKenzie et al. (2010) also showed that the ASD group had no trouble generating disablers for given conditionals. In terms of the model proposed here, the explanation of these phenomena is that people with ASD do not represent conditionals with a hidden exception-parameter, to which exceptions or disabling conditions can be attached in the manner described in section 2.1. In (Stenning & van Lambalgen, 2008, Chapter 9) we use neurophysiological data to further flesh out this explanation. In coarsest outline, the argument is this. A conditional 'if p then q' is represented by neurons n_p, n_q and an excitatory link from n_p to n_q. In the expanded form of the conditional, 'if p and *not-e* then q', *not-e* represents an inhibitory neuron i_e, which interrupts or at least decreases the flow of activation from n_p to n_q. The neuron i_e collects and sums the activations of the (neurons corresponding to the) exceptions connected to e. There are several reasons why inhibition is compromised in ASD (detailed in (Stenning & van Lambalgen, 2008, Chapter 9)), e.g. immature inhibitory interneurons, and a pruning process which selectively spares excitatory neurons. Any of these processes would impair i_e's ability to inhibit the flow of activation from n_p to n_q.

7. Conclusion: cohesion, coherence and reasoning

The suppression paradigm is one of the richest in the field, but this can be seen only if the task is not conceptualised as a tug-of-war between (correct) classical logic and various biases luring the reasoner to incorrect conclusions. What makes the task so interesting (and relevant to the study of ASD) is that it investigates various ways of integrating premises, and their effects upon reasoning. In other words, we interpret the task as an instance of achieving *coherence*, and its linguistic counterpart *cohesion*. Discourse cohesion refers to the linguistic devices, such is interclausal connectives, used to establish *local* coherence, e.g.

between sentences. The classic reference for cohesion is (Halliday & Hasan, 1976):

> Cohesion occurs where the *interpretation* of some element in the discourse is dependent on that of another. The one presupposes the other in the sense that it cannot be effectively decoded except for recourse to it. When this happens, a relation of cohesion is set up, and the two elements, the presupposing and the presupposed, are thereby potentially integrated in the text. (Halliday & Hasan, 1976, p. 4).

The authors identify *conjunctions* (in the linguistic sense) as a means to achieve cohesion. They note, however, that the logical conjunction *and* is the least cohesive of all conjunctions: the truth table definition treats the conjuncts as independent, hence destroys cohesion. A less trivial example is the 'adversative conjunction' *but*. For instance, a basic conditional 'if p then q' and an alternative conditional 'if r then q' indicating an alternative cause can be joined in one expression

$$\text{if } p \text{ then } q, \textit{ but } \text{if } r \text{ then } \textit{also } q$$

Here the clause starting with *but* is dependent on the main clause 'if p then q', hence the two clauses form a cohesive discourse, even though from the viewpoint of classical logic the two clauses are connected by the truth-functional conjunction. We have seen that subjects with ASD are capable of this relatively superficial type of coherence.

There exists a deeper form of discourse coherence, however, which refers to the overall event structure determined by a discourse, including its connections to general world knowledge. This *global* structure can be realised by relating events with respect to time, space, causality, goal-plan-action, and internal composition.[15] This global event structure will in general contain events (or event-variables) which are not verbalised but subserve coherence by facilitating long-distance connections. The hidden exception parameter is an example of such a variable, and we have seen that coherence suffers when it is impaired, or even absent. Although we have hardly discussed time, we believe that time is a domain in which central coherence and its impairment manifest themselves profoundly. We provide one illustration, which concerns what is for neurotypicals one of the most basic laws of change in time: the law of inertia, which says that[16]

(25) if a property P holds at time t, and nothing happens between times s, t with $t < s$ that affects P, then P holds at s.[17]

[15] An example of the latter is the presence or absence of a canonical goal of an event. 'Writing a letter' has a canonical goal, 'writing letters' doesn't.

[16] For simplicity time is represented by a linear order $<$.

[17] Note the analogy with Newton's First Law. A formalisation is given in (van Lambalgen & Hamm, 2004, Chapter 4); it will be clear that it has the form of the main conditional of the suppression task, as given in **??**.

If the theory developed in this chapter is correct, people with ASD cannot have the representation (25). To see that this a cognitive impairment, consider the following example of counterfactual reasoning.

The counterfactual task designed by Riggs and Peterson (2000) is a version of the false belief task 'with the beliefs stripped off'. A typical story plus counterfactual question goes like this (there were four such stories)

> John and Mary are in the kitchen. They have some chocolate. They put the chocolate in the fridge. Then John leaves the house to go and visit a friend. Mary wonders to herself, "What will I do today? I know," she says, "I'll bake a cake." She goes to the fridge and gets the chocolate. She makes a cake with some of the chocolate and puts the rest of it away in the cupboard. If Mary had not baked a cake, where would the chocolate be?

Neurotypical reasoning could go like this. Consider Mary's baking a cake an exceptional event. Before that event, the chocolate was in the fridge, and since no event is introduced as occurring after the 'bake' event, closed world reasoning and (25) entail that the chocolate remains in the drawer. A further application of closed world reasoning yields that if Mary hadn't baked a cake, nothing would have happened which could affect the chocolate. By the law of inertia it follows that the chocolate remains where it is.

Inertia and closed world reasoning allow one to take a global perspective on time; but how can a subject with ASD reason himself to a solution? At this point we must recall that roughly half of subjects with ASD endorses DA. Let p be 'Mary bakes a cake', and q: 'chocolate moves to drawer'. The counterfactual question 'If Mary had not baked a cake, where would the chocolate be?' must be answered on the basis of the premises '*not-p*; if p then q'. DA gives '*not-q*', which by closed world reasoning implies that the chocolate is in the fridge. If the subject does not have the law of inertia and does not endorse DA, the counterfactual question seems intractable. Grant, Riggs, and Boucher (2004) compared counterfactual reasoning of neurotypical children and children with ASD, using the materials of Riggs and Peterson (2000). The neurotypical group scored near-ceiling, in the ASD group 44% answered all four questions correctly. These results are consistent with the analysis just given.

References

Alvarez, A. L., & Booth, A. E. (2015). Preschoolers prefer to learn causal information. *Frontiers in Psychology*, *6*, 60. doi: 10.3389/fpsyg.2015.00060

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge: Cambridge University Press.

Baggio, G., van Lambalgen, M., & Hagoort, P. (2008). Computing and recomputing discourse models: An erp study. *Journal of Memory and Language*, *59*(1), 36–53. doi: 10.1016/j.jml.2008.02.005

Byrne, R. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31*, 61–83.

Congiu, S., Schlottmann, A., & Ray, E. (2010, Jan 01). Unimpaired perception of social and physical causality, but impaired perception of animacy in high functioning children with autism. *Journal of Autism and Developmental Disorders*, *40*(1), 39–53. doi: 10.1007/s10803-009-0824-2

Csibra, G., & Gergely, G. (2007). 'obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*, 60–78.

Declerck, R., & Reed, S. (2001). *Conditionals. a comprehensive empirical analysis*. Berlin/New York: Walter de Gruyter.

Diehl, J. J., Bennetto, L., & Young, E. C. (2006, Feb 01). Story recall and narrative coherence of high-functioning children with autism spectrum disorders. *Journal of Abnormal Child Psychology*, *34*(1), 83–98. Retrieved from `https://doi.org/10.1007/s10802-005-9003-x` doi: 10.1007/s10802-005-9003-x

Fillenbaum, S. (1978). How to do some things with if. In Cotton & Klatzky (Eds.), *Semantic functions in cognition.* Maweh, NJ: Lawrence Erlbaum.

Frye, D., Zelazo, P. D., & Burack, J. A. (1998). Cognitive complexity and control: I. theory of mind in typical and atypical development. *Current Directions in Psychological Science*, *7*(4), 116-121. doi: 10.1111/1467-8721.ep10774754

Grant, C. M., Riggs, K. J., & Boucher, J. (2004, Apr 01). Counterfactual and mental state reasoning in children with autism. *Journal of Autism and Developmental Disorders*, *34*(2), 177–188. doi: 10.1023/B:JADD.0000022608.57470.29

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

Hughes, C., & Russell, J. (1993). Autistic children's difficulty with disengagement from an object: its implications for theories of autism. *Developmental Psychology*, *29*, 498–510.

Loth, E., Gómez, J. C., & Happé, F. (2008). Event schemas in autism spectrum disorders: The role of theory of mind and weak central coherence. *Journal of Autism and Developmental Disorders*, *38*(3), 449–463.

Mayer, M. (1969). *Frog, where are you?* New York: Dial books for young readers.

McKenzie, R., Evans, J., & Handley, S. (2010, 3). Conditional reasoning in autism: Activation and integration of knowledge

and belief. *Developmental Psychology*, *46*(2), 391–403. doi: 10.1037/a0017412

McKenzie, R., Evans, J. S. B. T., & Handley, S. J. (2011). Autism and performance on the suppression task: Reasoning, context and complexity. *Thinking & Reasoning*, *17*(2), 182-196. doi: 10.1080/13546783.2010.549302

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255–258.

Pijnacker, J., Geurts, B., van Lambalgen, M., Buitelaar, J., Kan, C., & Hagoort, P. (2009). Conditional reasoning in high-functioning adults with autism. *Neuropsychologia*, *47*(3), 644–651.

Riggs, K., & Peterson, D. (2000). Counterfactual reasoning in preschool children: Mental state and causal inferences. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 87–100). Psychology Press.

Russell, J., Hala, S., & Hill, E. (2003). The automated windows task: the performance of preschool children, children with autism, and children with moderate learning difficulties. *Cognitive Development*, *18*(1), 111 - 137. doi: https://doi.org/10.1016/S0885-2014(02)00163-6

Stenning, K., & van Lambalgen, M. (2008). *Human Reasoning and Cognitive Science*. Cambridge, MA: MIT University Press.

Tager-Flusberg, H. (1995). 'Once upon a ribbit': Stories narrated by autistic children. *British Journal of Developmental Psychology*, *13*(1), 45–59.

van Lambalgen, M., & Hamm, F. (2004). *The proper treatment of events*. Oxford and Boston: Blackwell.

Vivanti, G., McCormick, C., Young, G. S., Abucayan, F., Hatt, N., Nadig, S., A.and Ozonoff, & Rogers, S. J. (2011). Intact and impaired mechanisms of action understanding in autism. *Developmental Psychology*, *47*(3), 41-56.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young childrens understanding of deception. *Cognition*, *13*, 103–128.

University of Edinburgh
E-mail address: k.stenning@ed.ac.uk

ILLC, University of Amsterdam
E-mail address: m.vanlambalgen@uva.nl