Logic and Psychology: A Couple of Case Studies

Torben Braüner

Roskilde University, Denmark

Abstract

In this conceptual paper we first make some general remarks on logic in real human reasoning, logic in mathematical reasoning and logic in philosophy (sections 1, 2, 3). We then move on and consider logic in psychology (Section 4) and we describe a couple of psychological studies where logical reasoning plays a decisive role. The psyhological studies in question involve what are called false-belief tests (Section 5) and tests where an experimental subject has to judge whether a syllogism is valid, despite inconsistent contextual information (Section 6). In Section 7 we make some remarks about further work.

Keywords

Logic and psychology, False-belief tests, Syllogistic reasoning, Bias in reasoning, Autism Spectrum Disorder (ASD)

1. Introduction: Logic in real human reasoning

Logic is usually taken to be the study of *correct* reasoning. Apparently, people have an ability to do logical reasoning, and to distinguish logically correct reasoning from incorrect.[1]

The history of logic goes back to the famous Greek philosopher Aristotle (384–322 BC).[2] Amongst many other achievements, Aristotle worked out systematic principles for correct reasoning, in particular, he identified and classified certain patterns of reasoning called syllogisms, having two premises and one conclusion, like the following.

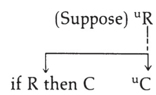
All men are humans
All humans have brains
——————————————————
Therefore, all men have brains

We say that a syllogism is valid if and only if the truth of the premises imply the truth of the conclusion, just by virtue of the logical form of the three statements. This is the case with the syllogism above. Thus, the content of the statements is irrelevant, in particular, the actual meaning of "men", "humans" and "have brains" does not matter, for example, the word "men" could be replaced by "politicians" without affecting validity. In this sense validity is a topic-neutral notion.

Since Aristotle's introduction of syllogisms, these patterns of reasoning have been a benchmark for reasoning studies in various different disciplines, including philosophical logic and psychology; see [2] for a comprehensive overview and meta-analysis of 12 different psychological theories of syllogistic reasoning.

Nowadays, the significance of logical reasoning in everyday life is witnessed by the contemporary critical thinking literature where logical notation is used to analyze actual human reasoning. In particular, the structure of informal arguments is described using what are called argument diagrams, thereby helping to judge whether an argument is valid or not. Example: A common reasoning pattern is *conditional reasoning*, that is, we often make an assumption ("for the sake of argument"), and then work out the consequences of the assumption in question. In the critical thinking textbook [3], page 122, the following notation is used for conditional reasoning.

> In general if an argument proceeds from supposition R to conclusion C and then concludes 'if R then C' we shall represent this process of conditionalisation in an argument diagram as follows,



Using such notation, the book [3] analyses many different sorts of arguments: Scientific arguments, arguments about God's existence, strategic reasoning about nuclear deterrence, and a number of others. The take-home message here is that applications of logic abound in real everyday human reasoning.

2. Logic in mathematical reasoning

Of course, a fantastic example of logical reasoning is the reasoning that takes place in mathematics, which is based on clear-cut logical reasoning principles. This is witnessed by formalisations, where logical proofs built according to the rules of a precisely defined proof-system can be used to represent—describe the structure of—actual mathematical proofs, carried out by real human mathematicians.

The idea of formalising mathematical proofs using such precisely defined proof-rules traces back (at least) to Gerhard Gentzen's work in the 1930s, cf. [4], page 74.

> We wish to set up a formalism that reflects as accurately as possible the actual logical reasoning involved in mathematical proofs.

To this end, Gentzen invented what are nowadays called natural deduction style proof systems for first-order logic. The goal of natural deduction systems to mimick actual mathematical reasoning has been repeated many times since then,[3] in particular, it is the cornerstone in the works of Dag Prawitz, see for example the classic book [5]. See [6] for a general introduction to natural deduction systems.

10th Workshop on Formal and Cognitive Reasoning (FCR-2024) at the 47th German Conference on Artificial Intelligence (KI-2024, September 23 - 27), Würzburg, Germany
torben@ruc.dk (T. Braüner)
http://akira.ruc.dk/~torben/ (T. Braüner)
© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

[1] Possible objection: People make errors in logical reasoning tasks, sometimes even systematically! Response: True, but people usually understand the correct solution when it is explained to them (with some effort, perhaps). See also Footnote 5.

[2] At least as far as logic in the Western world is concerned. There are other logic traditions, for example Indian logic, see [1].

[3] Comment from the author: It is definitely an arguable claim that natural deduction rules reflect actual mathematical reasoning step-by-step, but it is surprising that there does not seem to much published work trying to provide empirical evidence for the claim in the form of statistically valid experimental studies.

CEUR-WS.org/Vol-3763/paper1.pdf
CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

But how does a natural deduction system look? Well, a natural deduction system for ordinary propositional logic can be seen in Figure 1. There are two main ideas behind such a system:

The first idea is that there are two different kinds of rules for each logical connective: one to introduce it and one to eliminate it (rules are read from top to bottom). Introduction rules have names in the format $(...I...)$ and elimination rules have names in the format $(...E...)$.

The second idea behind a natural deduction system is that conditional reasoning is hardwired into the system, that is, at any stage in a deduction we can

1. make a new assumption,
2. work out its consequences,
3. and then discharge it (disregard it).

The definition of this discharge mechanism is a bit technical, but briefly put, discharged assumptions are indicated by putting parantheses $[\dots]$ around them, see the rules in Figure 1. Note that the rule $(\bot)$ formalize the standard mathematical proof technique called proof by contradiction: If it is assumed that ϕ is false, that is, $\neg\phi$ is true, and this assumption implies a contradiction, denoted $\bot$, then it can be concluded that ϕ is true.

Conditional reasoning is a very common reasoning pattern in mathematics as well as in informal reasoning (note the similarity between the rule $(\to I)$ in Figure 1 and the above notation for conditional reasoning taken from the critical thinking book [3]). Also, there is some experimental backing for the claim that natural deduction style reasoning is a mechanism underlying human thinking more generally, cf. what is called the mental logic school in the psychology of reasoning, which we describe in Subsection 4.1.

The natural deduction system for propositional logic has desirable mathematical properties: The system can be equipped with reduction (rewrite) rules that can remove detours when the introduction of a connective is followed by an elimination, see Figure 2 where $\tau \rightsquigarrow \theta$ means that the derivation τ is rewritten to the derivation θ. This leads to the following:

Definition 1. *A derivation is* normal *if no reductions can be applied.*

Theorem 1. *(Normalization) Any derivation can be rewritten to a normal derivation by repeated applications of reductions.*

Theorem 2. *(Subformula property) All formulas in a normal derivation, except some trivial cases, are subformulas of the end-formula or undischarged assumptions.*

The exact formulations and proofs of the above properties can be found in many different places, for example [5]. Nowadays, natural deduction systems are available in many different variants for many different logics.

$$\dfrac{\phi \qquad \psi}{\phi\wedge\psi}\,(\wedge I) \qquad\qquad \dfrac{\phi\wedge\psi}{\phi}\,(\wedge E1) \qquad \dfrac{\phi\wedge\psi}{\psi}\,(\wedge E2)$$

$$\dfrac{\begin{array}{c}[\phi]\\ \vdots\\ \psi\end{array}}{\phi\to\psi}\,(\to I) \qquad\qquad \dfrac{\phi\to\psi \qquad \phi}{\psi}\,(\to E)^{*}$$

$$\dfrac{\begin{array}{c}[\neg\phi]\\ \vdots\\ \bot\end{array}}{\phi}\,(\bot)^{\star}$$

$*$ This proof principle usually goes under the name *modus ponens*.
$\star$ Side-condition for technical reasons: ϕ is a propositional symbol ($\neg\phi$ abbreviates $\phi\to\bot$). This proof principle is usually called *proof by contradiction*.

Figure 1: Natural deduction rules for propositional logic

$$\dfrac{\dfrac{\begin{array}{c}\vdots\,\pi_1\\ \phi\end{array}\qquad\begin{array}{c}\vdots\,\pi_2\\ \psi\end{array}}{\phi\wedge\psi}\,(\wedge I)}{\phi}\,(\wedge E1) \qquad\rightsquigarrow\qquad \begin{array}{c}\vdots\,\pi_1\\ \phi\end{array}$$

$$\dfrac{\dfrac{\begin{array}{c}[\phi]\\ \vdots\,\pi_1\\ \psi\end{array}}{\phi\to\psi}\,(\to I)\qquad \begin{array}{c}\vdots\,\pi_2\\ \phi\end{array}}{\psi}\,(\to E) \qquad\rightsquigarrow\qquad \begin{array}{c}\vdots\,\pi_2\\ \phi\\ \vdots\,\pi_1\\ \psi\end{array}$$

Figure 2: Natural deduction reduction (rewrite) rules for propositional logic

3. Stepping stone: Logic from a philosophical point of view

Two main branches of logic can be distinguished, in terms of the subject as well as the way researchers identify themselves. A notable difference between the branches is that they have distinct success criterias regarding what a "good" proof-system is.

The first branch is model-theory: It is basically what we nowadays call semantic. Propositions and logical connectives are interpreted in terms of set-theoretic models, a notion which can be traced back to Alfred Tarski's definition of truth from the 1930s, [7]. That is, this branch takes as a cornerstone the relation between formulas and models defined by recursive truth-conditions, where models in many cases are thought of as representing a specific part of reality. In this way model-theory lends itself towards realism, that is, the philosophical view that reality exists independently of what human beings think or say about it. A model-theorist will typically ask whether a given proof-system is sound and complete wrt. a model-theoretic semantics (in propositional logic: a formula is provable if and only if it is a tautology).

The second branch is proof-theory: It is basically syntactic, but attempts to locate the meaning of a connective in the role it plays in logical rules. This has developed into a separate semantic paradign called proof-theoretic semantics, sometimes associated with Ludwig Wittgenstein's language games cf. his slogan 'meaning is use'. See the paper [8] for a presentation of proof-theoretic semantics, and note that that paper, page 9, points out that Wittgenstein's slogan 'meaning is use' should be understood in a normative way as 'meaning is correct use' to distinguish it from factual linguistic behaviour of speakers.

Now, a proof-theorist will typically ask whether a natural deduction system can be equipped with reduction rules such that a normalization result can be proved, where normal derivations satisfy a subformula property, as we saw earlier in Section 2. In the case of a sequent system, a proof-theorist will ask for a cut-elimination theorem (preferably proved in a syntactic, rather than semantic, way). Such a theorem says that for any derivation including applications of the cut rule (that is, lemmas), there exists a derivation of the same end-sequent, but without applications of cuts. In most sequent systems, cut-free derivations satisfy a subformula property. To quote the famous proof-theorist Jean-Yves Girard: "A logic without cut-elimination is like a car without an engine."[4]

The dichotomy between model-theory and proof-theory can be found in philosophical-, mathematical- as well as computational logic, even though it is not always explicitly articulated. See [8] for a presentation of a number of different semantic paradigms.

4. Logic in psychology

How are logic and psychology related, if at all related? There are basically two different claims:

The first claim is that psychology is relevant to logic. This claim has been rejected by Gottlob Frege (1848–1925) and many later logicians, who said that psychology is *descriptive* whereas logic is *normative*: How people *actually* reason is irrelevant to how they *should* reason.

The second claim is that logic is relevant to psychology. This claim is rejected by some psychologists, one reason being that people do not perform well in certain logical reasoning tests, for example what is called the Wason Card Task.[5] The rejection of the second claim has been criticized for several reasons, in particular for being based on a too narrow view of logic, see for example the book [12], which gives a detailed account of the relationship between logic and psychology, historically as well as more recent developments.

In the book [12], logic is used to explain the reasoning of experimental subjects in a number of different psychological reasoning tasks, thus, the book uses normative considerations when explaining actual human reasoning; which in the paper [13] is called normatively informed descriptive work. We shall later in the present paper corroborate the second claim by giving two case studies, also showing the relevance of logic to psychology.

4.1. Using logic in psychology, more concretely

In Section 3 we briefly described the two main branches in logic namely model-theory and proof-theory. These two branches correspond (roughly) to two schools in the psychology of reasoning (there are other schools, for example based on probability theory, but they are omitted here).

The model-theory branch of logic corresponds to the "mental models" school, according to which the cognitive mechanism underlying human reasoning is the construction of models. The structure of such a mental model is analogous to the structure of the situation it represents. This view is held by Philip Johnson-Laird and others, see [14].

On the other hand, **the proof-theory branch of logic corresponds to the "mental logic" school**, according to which the mechanism underlying human reasoning is not the construction of models, but rather the application of topic-neutral formal rules. To be more precise, according to this school, natural deduction style rules are somehow built into the human cognitive architecture. Thus, the starting point is here linguistic (syntactic) representations. This view has for example been held by Lance Rips, see the book [15] where he provides experimental support for the claim.

> ...a person faced with a task involving deduction attempts to carry it out through a series of steps that takes him or her from an initial description of the problem to its solution. These intermediate steps are licensed by mental inference rules, such as modus ponens, whose output people find intuitively obvious. ([15], p. x)

Note that modus ponens is the natural deduction rule $(\rightarrow E)$ in Figure 1. See also [16] which is a reproduction of some chapters from the book [15].

Both of the above schools are reflected in different psychological theories of syllogistic reasoning, see the account given in [2]. See also the discussion of syllogisms in [12], in particular, the book's discussion of the mental models theory of syllogistic reasoning.

5. Case study: Logical formalization of false-belief tasks

In this section, we describe a line of work where we use logic to investigate psychological reasoning tests called false-belief tasks. More precisely, the goal is to analyze and give logical formalizations of false-belief tasks using proof-systems for hybrid modal logic, and moreover, to use such logical analyses to explicate how individuals with Autism Spectrum Disorder (ASD)[6] reason in false-belief tasks.

A well-known example of a false-belief task is what is called the Smarties test. The following is one version of this test.

> A child is shown a Smarties tube where unbeknownst to the child the Smarties have

[4] There are dissenting voices, though, for example the paper [9] defends analytic cuts (where the cut formula is a subformula of the end-sequent) on the basis that the elimination of all cuts gives rise to various kinds of anomalies, like it is in [10] pointed out that there are first-order formulas whose derivations in cut-free systems are much larger than their derivations in natural deduction systems, which implicitly allow unrestricted cuts (in one case more than 10^{38} symbols compared to less than 3280 symbols).

[5] Johan van Benthem gave the following succint remark in connection with the Wason Card Task, cf. [11], page 77: "A psychologist, not very well-disposed toward logic, once confessed to me that despite all problems in short-term inferences like the Wason Card Task, there was also the undeniable fact that he had never met an experimental subject who did not understand the logical solution when it was explained to him, and then agreed that it was correct. Why should the latter slightly longer-term 'reflective fact' be considered less of a cognitive reality than the former?"

[6] Autism Spectrum Disorder is a psychiatric disorder with the following diagnostic criteria: 1. Persistent deficits in social communication and social interaction. 2. Restricted, repetitive patterns of behavior, interests, or activities. For details, see *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-V)*, published by the American Psychiatric Association.

> been replaced by pencils. The child is asked: "What do you think is inside the tube?" The child answers "Smarties!" The tube is then shown to contain pencils only. The child is then asked: "If your mother comes into the room, and we show this tube to her, what will she think is inside?"

It is well-known from experimental studies that most children above the age of four correctly say "Smarties" (thereby attributing a false belief to the mother) whereas younger children say "Pencils" (what they know is inside the tube). For children with autism, the cutoff age is higher than four years. This difference was observed already in [17] in connection with another false-belief task called the Sally-Anne task.

Passing the Smarties test involves taking the perspective of another agent namely the mother, and reasoning about what she believes. The child has to put himself/herself in the mother's shoes to get the answer right. Since the ability to take a different perspective is a precondition for figuring out the correct answer to the Smarties (and other) false-belief tasks, the fact that children with ASD have a higher cutoff age is taken by many researchers to support the hypothesis that there is a link between autism and a lack of what is called *theory of mind*, which is the ability to ascribe mental states, for example, beliefs, to oneself and to others. For a general formulation of the theory of mind deficit hypothesis of autism, see the book [18].

In [19] we formalized the Smarties task described above, as well as the Sally-Anne task, using a natural deduction proof system for hybrid modal logic. Hybrid (modal) logic is an appropriate tool to analyze the reasoning in these false-belief tasks since it can explicitly represent perspectives.

More formally, hybrid logic is an extension of ordinary modal logic allowing explicit reference to individual points in a Kripke model, where—as usual in modal logic—the points stand for times, possible worlds, persons, or something else. The extra expressive power is obtained by adding what are called *nominals*, which are propositional symbols of a new sort, each being true at exactly one point in the Kripke model. In the temporal case, this means that we can formalize statements like

> it is Christmas Eve 2023.

This statement is true at exactly one time, namely Christmas Eve 2023. Besides nominals, a new kind of modal operators called *satisfaction operators* are added, enabling the formulation of statements being true at a particular point, for example, again in the temporal case

> at Christmas Eve 2023, it is snowing

In general, if a is a nominal and ϕ is a formula, then a new formula $@_a\phi$ can be built, where $@_a$ is a satisfaction operator. A formula of the form $@_a\phi$ is called a *satisfaction statement*. The formula $@_a\phi$ expresses that the formula ϕ is true at one particular point, namely the point to which the nominal a refers. Thus, $@_a\phi$ can be used to formalize statements like the one above saying that it snows at a particular point in time, namely Christmas Eve 2023. See [20, 21] for more details on hybrid logic.

The hybrid-logical natural deduction system used in [19] to formalize the Smarties and Sally-Anne tasks stems from our book [22] and can be traced back to [23]. The system in question is obtained by extending the natural deduction system for propositional logic of Figure 1 with the rules in Figure 3 (in fact, the system of [22] also includes rules for modal operators, but they are omitted here since we do not need them for the fomalizations we are interested in).

As usual in natural deduction systems, this system has proof rules for respectively introducing and eliminating a connective, which in the case of the satisfaction operator are the rules $(@I)$ and $(@E)$, cf. Figure 3. These rules formalizes respectively the two informal arguments

> It is Christmas Eve 2023
> It is snowing
> ———————————————————
> At Christmas Eve 2023, it is snowing

> It is Christmas Eve 2023
> At Christmas Eve 2023, it is snowing
> ———————————————————
> It is snowing

The hybrid-logical natural deduction system includes a further rule that we need to formalize the Smarties and Sally-Anne tasks, namely the somewhat technical rule denoted $(Term)$, cf. Figure 3. This rule allows us to jump to a hypothetical time (or whatever the points stand for), do some reasoning, and then jump back to the present time again. The hypothetical time is the time referred to by the nominal discharged by the rule, indicated by $[a]$. This nominal might be called the point-of-view nominal. Recall from Section 2 that parantheses $[\ldots]$ around an assumption means that it is discharged. Intuitively, the side-condition on $(Term)$ marked with $\star$ says that statements whose truth-values do not depend on time can be moved in and out of the hypothetical reasoning delimited by the rule (the vertical line of dots).

The rule $(Name)$ in Figure 3 is needed to prove a completeness theorem, see [22], but we do not need this rule for the formalizations we are interested in in the present paper, so we shall not comment more on it.

The hybrid-logical proof system in Figure 3 allows us to formalize the Smarties task as described in the beginning of this section. To this end we let nominals stand for persons. Then the shift of perspective in the Smarties task can be formalized very directly as the derivation in Figure 4 where we make use of the following symbolizations

D	Deduces that …
B	Believes that …
p	There are Smarties inside the tube
a	The imagined mother

and where the axiom $D\phi \to B\phi$ embodies the principle that if a person (an agent) deduces p, then the person comes to believe p. The reader might recognize the instance of the rule $(\to E)$ that was exhibited in Figure 1 and also the instances of hybrid-logical rules from Figure 3.

Note that in Figure 4, the nominal a stands for the mothers perspective and the end-formula $@_aBp$ of the derivation says that the mother believes that there are Smarties inside the tube. Moreover, note that the shift to the mothers perspective is dealt with by the $(Term)$ rule.

The Smarties test is described in many places in the psychological literature, and passing the test is usually described as involving a perspective shift. Such descriptions are informal, but we here consider a fully formal account of the child's reasoning in terms of a formalization using a

$$\dfrac{a \qquad \phi}{@_a\phi}\,(@I) \qquad\qquad \dfrac{a \qquad @_a\phi}{\phi}\,(@E)$$

$$\dfrac{\phi_1 \quad \ldots \quad \phi_n \qquad \begin{matrix}[\phi_1]\ldots[\phi_n][a]\\ \vdots \\ \psi\end{matrix}}{\psi}\,(Term)^{\star} \qquad\qquad \dfrac{\begin{matrix}[a]\\ \vdots \\ \psi\end{matrix}}{\psi}\,(Name)^{\dagger}$$

$\star$ The formulas $\phi_1, \ldots, \phi_n$ and ψ are all satisfaction statements and there are no undischarged assumptions in the derivation of ψ besides the specified occurrences of $\phi_1, \ldots, \phi_n$ and a.
† The nominal a does not occur in ψ or in any undischarged assumptions other than the specified occurrences of a.

Figure 3: Natural deduction rules for hybrid logic

$$\dfrac{@_aDp \qquad \dfrac{[a] \qquad \dfrac{\dfrac{[a] \quad [@_aDp]}{Dp}\,(@E) \qquad \dfrac{}{Dp \to Bp}\,(Axiom)}{Bp}\,(\to E)}{@_aBp}\,(@I)}{@_aBp}\,(Term)$$

Figure 4: Formalization of the child's correct response in the Smarties task (both person and temporal shift versions)

hybrid-logical proof system. Thus, a child's reasoning when giving a correct answer is explained by logical principles embodied in the hybrid-logical proof system.

In the terminology of the paper [13], the above line of work constitutes normatively informed descriptive work since we use logic to explain the reasoning of experimental subjects in a concrete reasoning task, thereby corroborating the second claim at the beginning of Section 4 that logic is relevant to psychology.

5.1. A second version of the Smarties task

Now, in the version of the Smarties task described above, the child had to take the mother's perspective, but there is a second version of the Smarties task requiring a shift of temporal perspective, but no shift of perspective to another person. The second version of the task is obtained by replacing the second question

> "If your mother comes into the room and we show this tube to her, what will she think is inside?"

by the following

> "Before this tube was opened, what did you think was inside?"

See [24] for more on the temporal version of the Smarties task. Of course, the correct answer is "Smarties" in both cases.

According to [19], the two versions of the Smarties task have the same formalization, where the nominal a in the temporal version refers to the time when the first question is asked. Thus, when formalizing the temporal version, the nominals stand for times. It follows that in the temporal version, the end-formula $@_aBp$ of the derivation in Figure 4 says that at the time the first question is asked, the child believes that there are Smarties inside the tube.

By formalizing the two versions of the Smarties task, we have not only shown that we can use logic to explain the reasoning in the two tasks, but we have also disclosed that the two seemingly different tasks have exactly the same underlying logical structure. Thus, passing the two tests can be explained by exactly the same logical principles. Again, we have corroborated the claim of Section 4 that logic is relevant to psychology.

5.2. Second-order false-belief tasks

The line of work described above dealt with *first-order* false-belief tasks; they are psychological tests where the experimental subject must ascribe a false-belief to oneself or another person. In a second-order false-belief task, the subject must keep track of a second person's belief about a third person's belief—it thus requires understanding of the recursive character of mental states.

Much less is known about second-order false-belief understanding than its first-order variant, in particular when it comes to children with ASD; see [25]. In the papers [26, 27], we considered a hybrid-logical formalization of a second-order false-belief task namely a second-order version of the above-mentioned (first-order) Sally-Anne task. This formalization highlights the importance of recursion: It shows that second-order reasoning can be viewed as the recursive embedding of first-order reasoning about different agents. More concretely, the hybrid-logical proof formalizing the second-order Sally-Anne task can be viewed as the embedding of the formalization of the first-order Sally-Anne task into a larger proof structure, capturing the second-order reasoning. Thus, another level of nesting is added to the perspectival analysis.

The paper [27] includes a logical comparison of the four well-known second-order false-belief tasks that can be found in the literature, showing that they are logically distinct and can be classified across two dimensions of variation. The empirical significance of the task classification was investigated in [28], where responses (for 41 neurotypical children and 62 children with ASD) on the four second-order false-belief tasks were analyzed using a Latent Class Analysis, which is a statistical method allowing the discovery of patterns in data that were not hypothesized beforehand. We were particularly interested in patterns involving *combinations* of tasks, for example, it turned out that for children with ASD, the conditional probability of passing the second-order Sally-Anne task, given that what is called the ice-cream task is passed, is close to 100 percent, but the converse conditional probability is 59 percent (note that this is stronger than just the observation that more subjects gave correct answers to the Sally-Anne task than to the ice-cream task). The results of [28] were based on data collected by Irina Polyanskaya as part of her PhD on second-order false beliefs in children with ASD, [29].

Note that the line of work described above constitutes normatively informed descriptive work, cf. [13], in two different ways: i) Logic is used to explain the reasoning in a concrete reasoning task namely the second-order Sally-Anne task, and ii) a logical classification of reasoning tasks is used as the basis of a statistical analysis of data in a concrete empirical study.

6. Case study: Syllogistic reasoning with bias

In this section we describe an analysis of certain logical reasoning tasks where autistic individuals perform not *worse*, but *better*, than typically developing individuals. This is for example what is reported in the paper [30], which compares a person's level of autistic-like traits to the person's ability to do syllogistic reasoning, cf. Section 1. See also [31].

Some syllogisms are consistent with reality: *All birds have feathers. Robins are birds. Therefore robins have feathers*, but others are not: *All mammals walk. Whales are mammals. Therefore whales walk.* Both of these syllogisms are valid, in fact they have exactly the same logical structure. These two syllogisms are of the respective types of valid-believable and valid-unbelievable (this terminology should be self-explanatory). But there are also the types invalid-believable and invalid-unbelievable. An example syllogism of the invalid-believable type is: *All flowers need water. Roses need water. Therefore Roses are flowers.* An invalid-unbelievable syllogism with exactly the same structure is: *All insects need oxygen. Mice need oxygen. Therefore mice are insects.*

It is well-known, cf. for example [32], that whether or not a syllogism is valid is easier to detect for congruent syllogisms (valid-believable and invalid-unbelievable) than for incongruent ones (invalid-believable and valid-unbelievable), this being the case because the correct answer to incongruent syllogisms is inconsistent with reality.[7] Thus, prior knowledge of reality can affect the judgment of validity. However, it also turns out this reasoning bias is smaller for autistic-like persons than for others: The study [30] shows that there is a negative correlation between this bias and what is called the AQ-score[8], thus, the more autistic-like a person is, the better the person is to judge syllogisms without being affected by irrelevant prior knowledge of reality.

To be more specific, in the study [30], each experimental subject judged four congruent syllogisms and four incongruent ones. With a 1-point score for each correct judgement, this gave rise to a 0-4 scale for congruent syllogisms and 0-4 scale for incongruent ones. Belief bias was calculated by subtracting the score for incongruent syllogisms from that of congruent ones, resulting in a possible belief bias score between -4 and 4 points for each subject. The paper [30] reports a correlation at -0.39 ($p < 0.001$) between AQ and belief bias. The study [30] does not break the -4 to 4 bias scale down into two -2 to 2 subscales for respectively valid and invalid syllogisms, so this work cannot clarify whether the judgement of valid and invalid syllogisms separately correlates negatively with the AQ-score, that is, whether autistic-like persons are better to judge both valid and invalid syllogisms, or only one of the types.

Now, in our paper [35] we asked the following question: What does it precisely mean that an experimental subject can judge a syllogism without bias, that is, without involving irrelevant contextual information? We assume that the validity of syllogisms is defined in the usual manner using first-order models (the syntax and semantics here are the standard machinery of first-order logic, so definitions are left out). Using this machinery, we can define a mathematical function `valid` which maps syllogisms to truth-values, that is, elements in the set $\{0, 1\}$. For example, if S is one of the first two syllogisms mentioned above, then $\texttt{valid}(S) = 1$, and if S is one of the last two syllogisms mentioned, then $\texttt{valid}(S) = 0$. Formally, the variable S here stands for a triple with three formulas that constitute a given syllogism, namely the two premises and the conclusion. With this definition, the function `valid` formalizes the normatively correct judgment of syllogisms.

Now, a subject's judgment of a syllogism takes place in a specific context, that is, in a specific state of affairs namely the actual state of affairs, where for example the statement *Robins have feathers* is true, but the statement *Whales walk* is false. Such a state of affairs is formalized by a model in first-order logic. This means that a specific subject's judgment of syllogisms in a context can be modeled by a function `believable` similar to the function `valid`, but with an extra parameter namely a model, representing a context. Thus, the function `believable` maps a pair consisting of a syllogism and a model to a truth-value, and the requirement of context-independence can be formulated as

$$(1) \qquad \texttt{believable}(S, \mathcal{M}_1) = \texttt{believable}(S, \mathcal{M}_2)$$

for any syllogism S and any models $\mathcal{M}_1$ and $\mathcal{M}_2$. Thus, a subject's judgment of syllogisms is context-independent if and only if the corresponding `believable` function satisfies the above requirement.

A stronger requirement than context-independence is correctness, that is,

$$(2) \qquad \texttt{believable}(S, \mathcal{M}) = \texttt{valid}(S)$$

for any syllogism S and any model $\mathcal{M}$. In particular, one may expect that a `believable` function is correct in this sense if it corresponds to a subject with logical training, who for example figures out whether a syllogism is valid by carrying out a paper-and-pencil calculation. Note that correctness is a strictly stronger requirement than context-independence, for example, a `believable` function that always gives the incorrect answer would obviously not be correct, but it would be context-independent. Note also that the model $\mathcal{M}$ does not occur on the right-hand side of (2), which reflects that the notion of validity is topic-neutral, cf. Section 1.

Note that the stimulus-response pattern of one experimental subject is modeled by one `believable` function. According to the definition above, the domain of such functions is constituted by all pairs consisting of a syllogism together with a first-order model determining the truth values of the premises and the conclusion of the syllogism. There are countably many such pairs, but of course, this infinite domain can be cut down to a finite domain by restricting to a finite set of syllogisms, for example in relation to a concrete psychological experiment.

In a sense, a `believable` function "measures" how logically correct a subject's reasoning is, and moreover, a `believable` function allows us to formulate precisely what it means that an experimental subject can judge syllogisms without bias. Concludingly, in the case study described in this subsection, logic is obviously relevant since the reasoning task itself is of a logical nature, so it is again corroborated that logic is relevant to psychology, cf. Section 4. Moreover, logic is the basis for a mathematically precise definition of what it means to be logically correct, cf. the paper [35]. In the terminology of the paper [13], the above line of work

[7] Such a reasoning bias is even found in inferences carried out by large language models, cf. [33, 34].

[8] The Autism-Spectrum Quotient (AQ) is a self-report questionnaire that measures the level of autistic-like traits.

can be characterized as normatively informed descriptive work.

7. Concluding remarks

In the present paper we have discussed common logical structures in various reasoning tasks: As discussed in Subsection 5.1, and originally pointed out in [19], two seemingly different versions of the Smarties task have exactly the same underlying logical structure, as demonstrated by hybrid-logical formalizations. Similarly, as discussed in Subsection 5.2, in [27] it was demonstrated that four second-order false-belief tasks share a certain logical structure, but they are also distinct in a logically systematic way. Our paper [36] compares the syllogistic reasoning task described in Section 6 to other psychological tasks where people with ASD outperform typically developing people, namely two decision-making task from behavioral economics as well as a task from the heuristics and biases literature, and in that paper we identified common formal structures (and differences). In fact, common logical structures in reasoning tasks were discussed already in the book [12] where it was demonstrated that a false-belief task has a logical structure similar to the structure of certain other reasoning tasks.

Now, besides being interesting in their own right, such analyses might also explain empirical results: If two experiments make use of distinct reasoning tasks, but which have the same underlying logical structure, then one might expect similar empirical results, in which case the identity of the logical structures can be seen as an explanation of the similarity of the results. Following such a strategy, our paper [28] investigated the empirical significance of a logical classification of four second-order false-belief tasks, as discussed in Subsection 5.2. Not only can logical analyses in such cases explain existing empirical results; we believe that logical analyses might even motivate new empirical experiments.

This raises the following question: How broad is the range of reasoning tasks are susceptible to formal and logical analyses? For example, can visual reasoning tasks be included? What we here have in mind are visual pattern matching tests where individuals with ASD or autistic-like traits have been known for a while to show superior performance, as manifested in the Embedded Figure Test (EFT), where experimental subjects have to spot a simple shape within a more complex one. Thus, the EFT test measures the ability to disentangle information from a context. See the meta-analysis [37].

Also, the above kind of research questions might not only be asked at the psychological level (with no reference to biology), but such questions can also be asked at the neurobiological level, for example along the lines of the fMRI study [38], which investigated brain correlates of syllogistic reasoning—it was found that two different types of syllogistic reasoning activated respectively what are called a parietal system and a left hemisphere temporal system. See also the more recent editorial [39] of a selection of papers that investigates the interplay between cognitive theories of reasoning and neuroimaging studies.

Acknowledgements

The research described in Subsection 5.2 was supported by the VELUX FOUNDATION (project *Hybrid-Logical Proofs at Work in Cognitive Psychology*, VELUX 33305). Thanks to Patrick Blackburn and Irina Polyanskaya for collaboration in connection with the VELUX-project. Thanks to Aishwarya Ghosh and Sujata Ghosh for collaboration in connection with our work described in Section 6. Thanks to the reviewers for detailed comments: Most comments have been taken into account, but due to lack of time to prepare the final version, some comments will have to wait for future work.

References

[1] S. Sarukkai, M. Chakraborty (Eds.), Handbook of Logical Thought in India, Springer New Delhi, 2022.

[2] S. Khemlani, P. Johnson-Laird, Theories of the syllogism: A meta-analysis, Psychological Bulletin 138 (2012) 427–457.

[3] A. Fisher, The logic of real arguments, second edition, Cambridge University Press, 2004.

[4] G. Gentzen, Investigations into logical deduction, in: M. Szabo (Ed.), The Collected Papers of Gerhard Gentzen, North-Holland Publishing Company, 1969, pp. 68–131.

[5] D. Prawitz, Natural Deduction. A Proof-Theoretical Study, Almqvist and Wiksell, Stockholm, 1965.

[6] A. Troelstra, H. Schwichtenberg, Basic Proof Theory, volume 43 of *Cambridge Tracts in Theoretical Computer Science*, Cambridge University Press, 1996.

[7] A. Tarski, The concept of truth in formalized languages, in: Logic, Semantics, Metamathematics, Oxford University Press, 1956, pp. 152–278.

[8] H. Wansing, The idea of a proof-theoretic semantics and the meaning of the logical operations, Studia Logica 64 (2000) 3–20.

[9] M. D'Agostino, M. Mondadori, The taming of the cut. Classical refutations with analytical cut, Journal of Logic and Computation 4 (1994) 285–319.

[10] G. Boolos, Don't eliminate cut, Journal of Philosophical Logic 13 (1984) 373–378.

[11] J. van Benthem, Logic and reasoning: Do the facts matter?, Studia Logica 88 (2008) 67–84.

[12] K. Stenning, M. van Lambalgen, Human Reasoning and Cognitive Science, MIT Press, 2008.

[13] C. Strößner, Hybrid frameworks of reasoning: Normative-descriptive interplay, in: Proceedings of the 45th Annual Meeting of the Cognitive Science Society, 2023, pp. 1153–1160. Available at https://escholarship.org/uc/item/58j925zr.

[14] P. Johnson-Laird, Mental models and deductive reasoning, in: J. Adler, L. Rips (Eds.), Reasoning: Studies of Human Inference and Its Foundations, Cambridge University Press, 2008, pp. 206–222.

[15] L. Rips, The Psychology of Proof: Deductive Reasoning in Human Thinking, MIT Press, 1994.

[16] L. Rips, Logical approaches to human deductive reasoning, in: J. Adler, L. Rips (Eds.), Reasoning: Studies of Human Inference and Its Foundations, Cambridge University Press, 2008, pp. 187–205.

[17] S. Baron-Cohen, A. Leslie, U. Frith, Does the autistic child have a 'theory of mind'?, Cognition 21 (1985) 37–46.

[18] S. Baron-Cohen, Mindblindness: An Essay on Autism and Theory of Mind, MIT Press, 1995.

[19] T. Braüner, Hybrid-logical reasoning in the Smarties

and Sally-Anne tasks, Journal of Logic, Language and Information 23 (2014) 415–439.

[20] T. Braüner, Hybrid logic, in: E. Zalta (Ed.), The Stanford Encyclopedia of Philosophy, Stanford University, 2021. On-line encyclopedia article available at http://plato.stanford.edu/entries/logic-hybrid.

[21] C. Areces, B. ten Cate, Hybrid logics, in: P. Blackburn, J. van Benthem, F. Wolter (Eds.), Handbook of Modal Logic, Elsevier, 2007, pp. 821–868.

[22] T. Braüner, Hybrid Logic and its Proof-Theory, volume 37 of *Applied Logic Series*, Springer, 2011.

[23] J. Seligman, The logic of correct description, in: M. de Rijke (Ed.), Advances in Intensional Logic, volume 7 of *Applied Logic Series*, Kluwer, 1997, pp. 107 – 135.

[24] A. Gopnik, J. Astington, Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction, Child Development 59 (1988) 26–37.

[25] S. Miller, Theory of mind: Beyond the preschool years, Psychology Press, 2012.

[26] T. Braüner, P. Blackburn, I. Polyanskaya, Recursive belief manipulation and second-order false-beliefs, in: Proceedings of the 38th Annual Meeting of the Cognitive Science Society, Philadelphia, Pennsylvania, USA: Cognitive Science Society, 2016, pp. 2579–2584.

[27] T. Braüner, P. Blackburn, I. Polyanskaya, Second-order false-belief tasks: Analysis and formalization, in: Proceedings of Workshop on Logic, Language, Information and Computation (WoLLIC 2016), volume 9803 of *Lecture Notes in Computer Science*, Springer-Verlag, 2016, pp. 125–144.

[28] T. Braüner, P. Blackburn, I. Polyanskaya, Being deceived: Information asymmetry in second-order false belief tasks, Topics in Cognitive Science 12 (2020) 504–534.

[29] I. Polyanskaya, Second-Order False Belief Reasoning by Children with Autism: A correlation and training study, Ph.D. thesis, Department of People and Technology, Roskilde University, Denmark, March 2019.

[30] M. Lewton, C. Ashwin, M. Brosnan, Syllogistic reasoning reveals reduced bias in people with higher autistic-like traits from the general population, Autism 23 (2019) 1311–1321.

[31] M. Lewton, Reasoning Behaviour Across the Autism and Psychosis Continuums, Lambert Academic Publishing, 2018.

[32] K. Klauer, J. Musch, B. Naumer, On belief bias in syllogistic reasoning, Psychological Review 107 (2000) 852–884.

[33] I. Dasgupta, A. K. Lampinen, S. C. Y. Chan, H. R. Sheahan, A. Creswell, D. Kumaran, J. L. McClelland, F. Hill, Language models show human-like content effects on reasoning tasks, 2024. URL: https://arxiv.org/abs/2207.07051. arXiv:2207.07051.

[34] R. Ando, T. Morishita, H. Abe, K. Mineshima, M. Okada, Evaluating large language models with neubaroco: Syllogistic reasoning ability and human-like biases, 2023. URL: https://arxiv.org/abs/2306.12567. arXiv:2306.12567.

[35] T. Braüner, A. Ghosh, S. Ghosh, Understanding responses of individuals with ASD in syllogistic and decision-making tasks: A formal study, in: Proceedings of 2nd International Workshop on Cognition: Interdisciplinary Foundations, Models and Applications, volume 12524 of *Lecture Notes in Computer Science*, Springer, 2021, pp. 118–128.

[36] T. Braüner, A. Ghosh, S. Ghosh, Understanding responses of people with ASD in diverse reasoning tasks: A formal study, Cognitive Processing (2024). To appear.

[37] S. Cribb, M. Olaithe, R. Di Lorenzo, P. Dunlop, M. Maybery, Embedded figures test performance in the broader autism phenotype: A meta-analysis, Journal of Autism and Developmental Disorders 46 (2016) 2924–2439.

[38] V. Goel, C. Buchel, C. Frith, R. Dolan, Dissociation of mechanisms underlying syllogistic reasoning, NeuroImage 12 (2000) 504–514.

[39] V. Goel, G. Navarrete, I. Noveck, J. Prado, Editorial: The reasoning brain: The interplay between cognitive neuroscience and theories of reasoning, Frontiers in Human Neuroscience 10 (2017) 1–2.