Postprint (please cite the published version)

Vieira, C., Mariz, D., Hoki, L. (2023). A Routine to Develop Inferencing Skills in Primary School Children. In: Mancenido-Bolaños, M.A.V., Alvarez-Abarejo, C.J.P., Marquez, L.P. (eds) Cultivating Reasonableness in Education. Integrated Science, vol 17. Springer, Singapore. https://doi.org/10.1007/978-981-99-4198-8_6

A Routine to Develop Inferencing Skills in Primary School Children

Celso Vieira, Débora Mariz, and Ligea Hoki[1]

Abstract

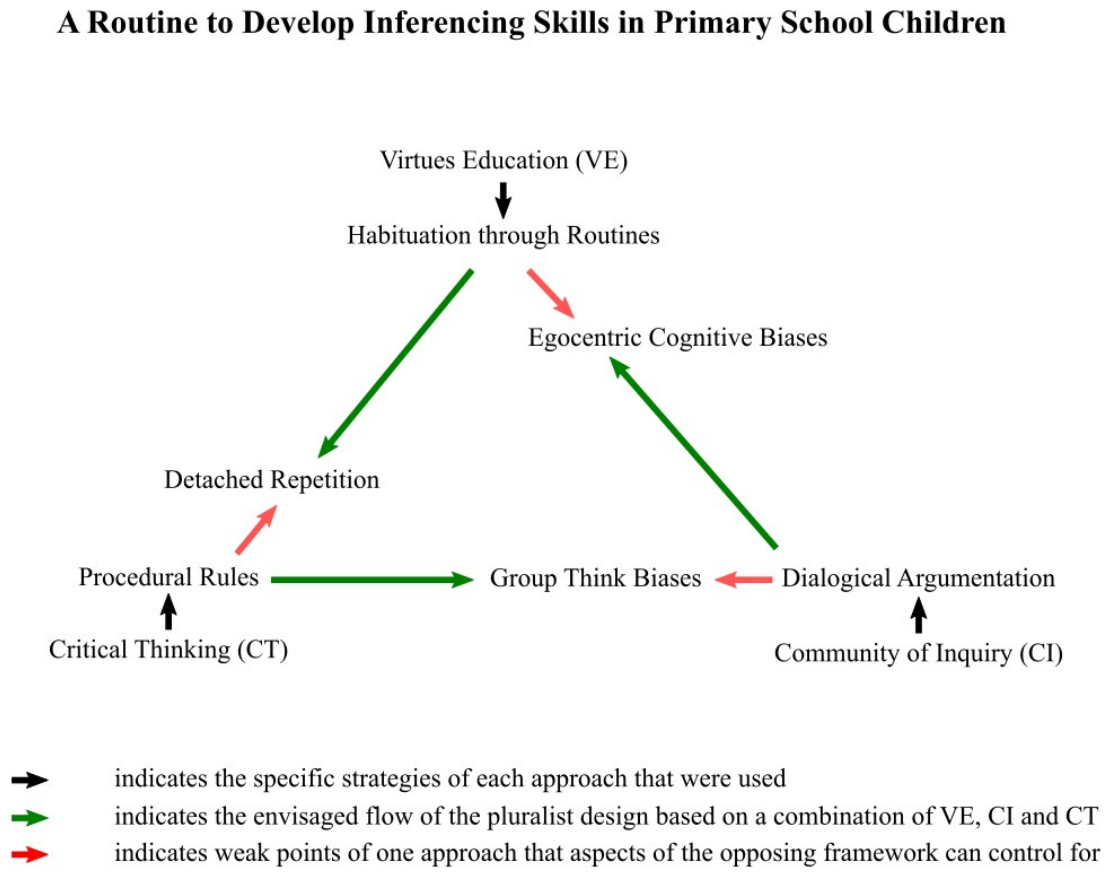
The chapter presents the prototyping of a thinking routine designed to foster good inference habits in children ages 6 to 11. The prototyping was developed at Ninho, an educational project for children from underprivileged households in Brazil. The thinking routines by Ritchhart and colleagues (2006) served as our starting point. Following a Virtue Education (VE) approach, we supposed that the repeated application would conduce to habituation. In addition, to increase peer-to-peer interactions, the teacher applying the routines worked as a facilitator in a Community of Inquiry (CI). After six months of application, the results were partially successful. We identified that the repeated exposition to the magic question “What makes you say that?” made children more aware of the evidence supporting their assumptions. Furthermore, the interactions between peers made them recognize different perspectives. However, we also identified some shortcomings. Most of them seemed to arise from cognitive biases. At the individual level, belief perseverance kept children too attached to their first assumptions. They preferred self-serving rationalizations instead of accepting a counter-argument. At the group level, we identified problems of social contagion such as information cascades. The effects of a first opinion rhetorically voiced were hard to efface. Since none of the steps on the previous routines addressed these biases, we prototyped a routine to start filling this gap. Following Critical Thinking (CT) theorists, we added a step of structured instruction concerning one specific reasoning technique - the inference to the best explanation. Moreover, the recommendation of cognitive psychologists motivated the inclusion of some extra features to avoid groupthink biases. For instance, a visual table that juxtaposes contrasting arguments should facilitate comparative evaluations. Also, a star-based evaluation scale should help different individuals to discuss their assessments based on a common ground. As described in the end, the prototype with elements from VE, CI and CT presented promising results.

Keywords

1***
Celso Vieira (RUB) Celso.DeOliveiraVieira@rub.com

Intellectual Virtues, Critical Thinking, Community of Inquiry, Inference to the Best Explanation, Philosophy for Children, Thinking Routines.

Visual Abstract



"Ninguém pode dizer a palavra verdadeira sozinho, ou dizê-la para os outros, num ato de prescrição, com o qual rouba a palavra aos demais."

"No one can say the true word alone or say it for the others in a prescribing act that steals the word from the others."

Paulo Freire

Introduction

This chapter presents the implementation of a thinking routine to develop in primary school children the ability to make inferences to the best explanation. Our general aim was to start developing Critical Thinking abilities in the pupils of the Ninho

Educational project. As described in section 1, we started with a weak pluralist approach. It combined the features of education for the intellectual virtues and the community of inquiry into a thinking routine based on Project Zero's 'see, think, and wonder'. Despite some positive results, our qualitative analysis revealed some shortcomings. Many of the problems seemed to be traceable to the occurrence of cognitive biases. At the individual level, children remained attached to an egocentric perspective while problems of social contagion became evident in the group dynamics.

Section 2 describes the two guiding strategies used to avoid the previously identified shortcomings. We followed the suggestion of some Critical Thinking theorists and added a step of structured instruction. Since the egocentric perspective seemed to be a problem of unaware inference, we opted to make the structure of the inference to the best explanation visible. Moreover, following the recommendation of cognitive psychologists, we added some extra features to control for the groupthink biases. A visual table that juxtaposes contrasting arguments should facilitate comparative evaluations. Finally, using a common evaluation scale should allow different individuals to discuss their assessments with less noise.

The implementation occurred within the context of Ninho, an educational project aiming to provide a top-notch education for children from low-income households in the region of Lagoa Santa/ MG in Brazil. The project has two phases. At first, Ninho provided scholarships for poor children to attend the best private school in the region. The gap between public and private education in Brazil demanded a significant amount of support for the children to keep up with their duties. Thus, they spent the counter-shift at Vila Ninho, the physical facility of the project. There, children had reinforcement classes, received help with the homework, and entertained more exploratory activities. In this context, we offered a series of routines to foster good habits of thinking. In 2022, the project entered a new phase and became a full-time school.

Three groups of pupils participated in the activities: 10 pupils in the 2nd grade, 10 pupils in the 3rd grade, and 6 pupils in the 5th grade (ages 6, 7, and 9 years old, respectively). No group had previous experience with philosophy classes. Each session of the activities lasted one hour per week during the whole year of 2021. The activities took place in the classroom, each grade separately.

In the first six months, we applied the thinking routines suggested by Project Zero, following the course 'Critical Thinking through Art' described in section 1.2 below. In the next six months, we prototyped a routine for the inference to the best explanation. The three of us participated as facilitators, but the activity had a hybrid setup. Débora was present in the classroom and acted as the main facilitator. Ligea and Celso joined the sessions remotely. The children could see and interact with the facilitators through a TV monitor.

The research followed a qualitative method of ethnographic observation [1]. Some of the sessions were recorded by video. Ligea took the field notes and made the observation protocols for the contribution of each pupil in every session. Celso was responsible for the analytical memos.

The ethnographic method suits our object of investigation, namely, the tacit inferential practices of children. The study design with recursive data collection and analysis allowed for ongoing prototyping of the routine. The analysis of the first set of

routines revealed how the pupils' unawareness of their own inferential process led to failures seeming to stem from associative memory. After consulting the literature on critical thinking [6, 7, 9] and cognitive psychology [28, 31, 32, 35], we developed a first prototype that should address the pupils' unawareness and biases. Repeated applications elicited more cases for recursive analyses. These were used to improve the prototype. Both stages are described in the following. Section 1 conveys our starting theoretical assumptions and how the data collection suggested a reformulation. Section 2 presents the final design.

1 The first application

1.1 The conceptual framework

Three pedagogical priority claims guided our decision to choose thinking routines as suitable activities to foster critical thinking at Ninho.

- The Community of Inquiry Claim (CI) circumscribes reasonableness. It claims that to foster good thinking in children priority should be given to the constitution of a communal structure in which the participants feel safe, supported, and motivated to engage in dialogic discussions with their peers. Participation in this community develops truth-conducive inquiry strategies [2].

- The Virtue Education Claim (VE) circumscribes a reasonable disposition. It claims that to foster good thinking in children priority should be given to nourishing proper intellectual disposition [3, 4]. Intellectual disposition includes the ability to reason, the sensibility to judge when such reasoning applies, and an inclination to be reasonable [5].

- The Critical Thinking Claim (CT) circumscribes reasoning. It claims that to foster good thinking in children priority should be given to the structured instruction of the techniques of good reasoning [6, 7]. These techniques consist of the principles of formal and informal logic.

These are priority claims, not exclusivity ones. Thus, each position has a different stance on what should be the prior concern of education without necessarily excluding the content of the other claims. If so, defenders of one position can reckon secondary roles to the circumscription of the others. For instance, CI enthusiasts are not claiming that we should *only* care about the community of inquiry without addressing disposition and reasoning techniques. Moreover, there is an overlap among these claims.[2] This justifies our use of the vague expression 'good thinking' as a shared goal for all three cases.

The use of the circumscriptions reasoning, reasonable and reasonableness also tries to do justice to the agreement among the views. It should be clear that they are not attached to each of the claims. For instance, one can foster reasoning without endorsing CT. Reasoning is defined as thinking constrained by inferential norms that one deems to be appropriate [8]. A reasonable disposition refers to the mental acts that make an agent act and react in a reasonable way [4]. Reasonableness encompasses the social aspect of being able to be reasoned with [2]. In the

2 The overlapping can also generate confusion. Both CI and VE use 'critical thinking' in a broader way than the one defined in CT above. In the following, we use it as in CT.

following, while talking about the approaches, we will refer to the claims CI, VE and CT, but while talking about the object, we can rely on the three Rs without committing to the claims.

Disagreements

The disagreements between the claims become more evident after noticing the difference in scope among them.

CT, the more restricted approach, assumes that structured instruction concerning the techniques of good reasoning is sufficient to develop good thinking. From a CT perspective, environment and disposition might help, but they are not necessary and certainly not sufficient for developing the competencies of good reasoning.[3]

CI and VE both argue for a wider scope than CT. VE is agent-centered. Thus, it focuses on cultivating dispositions instead of teaching techniques alone.[4] The community-centered CI is even wider and focuses on an environment in which several agents interact. VE and CI recognize the importance of reasoning techniques but also point out their insufficiency.[5] VE enthusiasts think that the direct instruction of the techniques of logical reasoning neglects the difference between having an ability and acquiring a disposition. For the CI enthusiasts, CT dismisses the role of peers, context sensibility and the affective aspects of pupils that are central to learning.[6]

The disagreement between CI and VE appears in the difference between top-bottom and bottom-up strategies.[7] VE defends an exemplarist top-down approach in which the students should have contact with virtuous reasoners to admire and emulate. CI, on the other hand, is peer-centered. The teacher plays the role of a facilitator in emphasizing the good practices in a dialogic discussion that should occur among the participants.

1.2 The first applications, soft-pluralism

3 See 7. Also, for Siegel [9] the virtuous intellect is not necessarily rational. Thus, in addition to intellectual virtues, one must learn thinking techniques. However, he concedes that VE provides better descriptions of the virtues that may help students.

4 See the Responsibilist approach to Intellectual Virtues [4]. For Baehr [10, p. 23], virtues contribute to becoming a better person while cognitive abilities do not.

5 Baehr [11, p. 258] admits that VE educators should focus not only on virtues but also on techniques. Bevan [12] and Batally [13] also argue for a joint approach. However, they all point out that the techniques are insufficient. We will argue for the necessity of a joint approach in a much stronger way.

6 Concerning CT, the materials of P4C include training on logic but, again, the focus is on peer-guided activities leading to the discovery and development of rules of deduction and avoidance of fallacies.

7 Again, there is an agreement. Baehr [3, Chapters 20 and 21] suggests creating the ideal classroom and environment for the development of virtues while reasonableness in CI is also dispositional.

The recognition of the overlap among the above mentioned positions convinced us that we could find a conciliatory activity that encompasses most of the positive aspects of the three. We call this first approach a soft pluralism because, at first, we did not think it would be necessary to make sure that each of the circumscriptions (the three Rs) was being addressed in a very well-defined way.

Based on the tension between VE+CI vs CT, we conceived an opposition that served as our starting guideline:

Exercise: the repetitive exercise of a competence stripped of any (or most) elements that are external or peripheral to that competence. Ex. Exercises on truth tables [7, p. 136].

Habituation: the repetitive exercise of a competence in an artificial scenario with external and peripheral elements. These elements emulate some of the features of real-life occasions in which the use of the competence will be valuable. The scenario tries to capture the social aspects and the role of peers.

We believe that detached practice, as in a series of exercises of formal logic, may be partially responsible for a backlash of a restricted approach to critical thinking by VE and CI. Moreover, the thicker conception of practice in habituation suits the thicker conception of good reasoning suitable to agent-based and community-centered approaches. Good reasoning should go beyond the hability to become a disposition and guide social interactions.

The willingness to contemplate a thicker conception of good reasoning led us to the 'making thinking visible' approach and its application through Project Zero's 'thinking routines' [14]. More specifically, we decided to start the implementation following the model of the course 'Critical Thinking through Art' (National Gallery of Arts).[8] The activities usually start with careful observation of a work of art followed by different activities designed to develop good thinking habits by following some ordered number of steps.

Take the routine 'See, think and wonder' as an example. The children are asked to observe a painting carefully, describe what they see, and tell what the previous steps made them think. After that, they take the whole experience a step further through wondering. In this step, the pupils provide the reasons that justify their wonder. The facilitator then uses the so-called magic question 'What makes you say that?' to prompt the pupils to become aware of their reasons.

The magic question nudges the pupils to expose and become aware of what may be grounding their thinking but does not make them engage in dialogic argumentation. To increase this type of peer interaction, the teacher assumed the role of a facilitator in a community of inquiry. Whenever conflicting interpretations arose, the children were encouraged to give evidence-based reasons for their views. They were also encouraged to appraise one another's reasons [15]. Thus,

Facilitator: You say that there is a wolf in the painting, did anyone say that it was something different?

Pupil: Yes, a dog.

8 The 'critical thinking' in the title is closer to the wider conception of critical thinking in CI and VE than that of CT defined above. For an evaluation of the program, see 16.

Facilitator: And what makes you think that it is a wolf and not a dog?

First results

After six months of application, a qualitative analysis of the notes and protocols of the routines revealed clear positive results. The aesthetic-related gains in observation[9], sensibilization[10] and justification[11] will not be topics of our reflection. Concerning reasoning, we could confirm that the strategy of making the thinking visible was responsible for much improvement in reflective thinking measured by the awareness of one's own assumptions.[12]

At first, the pupils were so eager to jump to conclusions that they did not differentiate between description (seeing), interpretation (thinking) and extrapolation (wondering). The facilitator almost naturally assumed the role of making these differences salient through questions such as 'Can you provide evidence?' or 'Where are you taking this information from?' Throughout the sessions, the pupils became more competent in stepping back and identifying the sources of their conclusions. However, after the first six months, they still had a hard time evaluating the degree of certainty of information arising from these different sources. For instance, they could not distinguish the likelihood of a conclusion based on observation from a fanciful association of ideas.

One extreme case may provide an illustration.[13] The children saw the film Luca in one of Ninho's movie afternoon sessions. The film features a half-fish, half-human 'hybrid' character. The following week, the children saw Gauguin's *Maria* in the thinking routine. It did not take long until one of the students suggested that the baby on the mother's shoulder was a 'hybrid'. Many others agreed.

Association of ideas is a source of creativity, and we want to be careful not to hinder children's imagination.[14] However, it is part of the required awareness for good reasoning that they assess the sources of their conclusions. During the first

9 Children became better at observation. In the beginning, the descriptions of what they saw were empty such as 'people', and 'trees'. We adopted a fractal method to push further descriptions. They name three elements in the painting. Then they had to give three adjectives for each element. Only after that, they were invited to say what they think concerning the described aspects.

10 There was also a visible gain in the familiarity with artists and artworks, with children being able to identify by style the paintings by Tarsila do Amaral or Van Gogh.

11 The children became more interested in the paintings, but there is much work to do. They usually rely on vague assessments such as 'I like it'. Thus, there is also room for development in the aesthetic realm.

12 See 18, p. 26.

13 Extreme cases offer good illustrations but their intensity should not be taken as representative.

14 Lipman et al [19, p. 62-64] identify the thin line educators have to walk in. Children are not good at reasoning, and their creativity does not survive schooling.

semester, we tried to raise their self-critical awareness relying on the suggestions by VE and CI.

1.3 Shortcomings in VE

Children are intuitive thinkers. Thus, one of the main steps in their philosophical education involves overcoming the egocentric perspective characterized by an unreflected certainty that theirs is the only view of the world [17, p. 301].[15] The information coming from the egocentric perspective plays a significant role in confirmation, egocentric, hindsight, and overconfidence biases. These, together with the availability heuristic, share the common feature of retrieving, selecting or even forging information that fits one's worldview.[16] In our experience, this type of behavior deeply influenced the children during the routines. For instance, in the face of the evidence that the animal in Kirchner's *View of Basel and the Rhine* could not be a horse because it was too small in relation to the humans, a pupil preferred to come up with a justification that it might be a little colt instead of changing his anchored view.

Since biases are so manifest in adults, it is not surprising that they occur even to a greater extent in children. We adopted two VE strategies in an attempt to avoid them.

Narrative examples: Following the exemplarist approach, defenders of VE suggest that an efficient way to elicit virtuous behavior is by telling narratives of virtuous agents in action [20]. In our applied context, we added stories of how the artists were constantly trying to improve their views to the point of abandoning strategies that were successful in their previous work. For instance, in a routine based on the *Great Wave*, the facilitator would tell how Hokusai spent his whole life trying to perfect his mastery. He even changed his name to mark these developments.[17]

Theoretical reflection: VE enthusiasts recommend including some theoretic discussion about what it is to have a particular intellectual virtue [3, p. 306]. Accordingly, we added some sessions in which we explained, discussed and praised intellectual humility. Some of the sessions were before the routines, others on separate days.

Our qualitative analysis in the first six months of application did not identify any effect on the egocentric selection of information. Narrative examples and theoretical reflection did not seem to diminish the prevalent biases.[18] We found a possible explanation for the failure in previous theoretic objections to the limitations of VE concerning biases. To put it simply, experts do acquire better intuitive judgments through habituation. However, experts also are susceptible to biases [21, p. 63-65].

15 Children also fail to notice that most of the contents of their views stem from inputs by adults through several cultural sources.

16 The literature on biases is extensive, see 25 for a general account.

17 Hokusai even wrote on the *Great Wave* that he was changing his name, thus the children could see the evidence.

18 This is a limited application both in time and number of participants. So we are not claiming that they do not work. We just did not see an effect.

Thus, even if a dispositional education provides a better judgment, the improvement will not necessarily work against biases.[19]

The behavior of children in our restricted experience confirmed this line of reasoning. During the theoretical reflection, the pupils could understand intellectual humility as the willingness to change their position in light of compelling counter-evidence or argument.[20] However, in practice, they remained prone to engage in fanciful rationalizations in order to keep their initial position instead of changing it.

The shortcoming also fits the CT critique of VE's ostensive pedagogy. VE is agent-based and not act-based. Thus, it avoids breaking down procedural recommendations to action, which it deems to be simplistic. Instead, agents will become virtuous by modeling virtuous agents [22, p. 154, 23: 15, 24: 17-18]. Once that is accepted, there is no specific set of techniques for good reasoning. This is criticized by CT enthusiasts [7, p. 132]. Exemplarism expects too much of novices. They must recognize what makes the behavior of a virtuous agent virtuous and figure out by themselves how to operate in difficult situations. Moreover, even if the novice acquires such a disposition, they will not be able to provide a conscious justification for their chosen behavior. In our experience, the opacity of acquiring a disposition also contrasts with making thinking visible, which proved to be one of the most useful features of the routines.

CT, on the other hand, is act-based. Accordingly, the CT teacher can provide structured instruction via a step-by-step procedure to evaluate arguments and a non-personal way to justify why the procedures work.[21] Psychologists studying bias in the judgment of experts agree that acquiring expertise by practice works. However, experts are still subject to biases which procedural rules and guidelines can reduce [28, chap. 8]. If so, procedural guidance plays an important role even when acquiring and cultivating good habits work.

Shortcomings in CI

Due to its social nature, CI offers a solution against the tendency to select only the information that confirms an egocentric perspective. A growing body of evidence corroborates that CI increases the force of arguments due to the consideration of opposing views [29, 30]. In our applications, the facilitator prompted pupils to appreciate their conflicting views. This peer interaction should lead them towards a relativistic view in which one recognizes that there are other ways to see the world [17, 8].[22] The goal, however, should lie in a step forward. From an intersubjective perspective, these different views are assessed collectively.

Psychological research confirms that, under certain conditions, the joint deliberation in group decision-making decreases confirmation biases because the participants

19 See 26. See 27 for an attempt to defend VE.

20 One difficulty that we had to overcome was children's tendency to mix humility and intellectual humility together.

21 As Kotzee et al [7, p. 136] argue, even if one is going to teach by exemplar narratives, it should be necessary to tell how the virtuous agent acted and why it is justified.

22 However, it is not so clear that the pupils are really open to going beyond the simple recognition of different views and are willing to change their minds. See, for instance, the identification that the students cannot build an argument for the opposing views in 33.

have to face the objections of others. However, the 'under certain conditions' apposition is important. In natural setups, spontaneous group discussion also generates an ideal environment for the manifestation of other biases, namely, those related to social influence.

In our application, CI did increase the recognition of other points of view. However, we also identified that the dynamic was susceptible to biases associated with groupthink [31]. More specifically, our qualitative analysis suggested that the group opinion depended more on information cascades than on the plausibility of a view.

A cascade in cognitive psychology refers to a biased information chain in which an arbitrary first opinion or evaluation generates a contagious trend that influences further ones [32]. Popularity proves to be so socially self-reinforcing that the effects of the first evaluations become almost impossible to efface. Evidence that the cascade relies on arbitrary factors comes from the fact that similar groups with similar information will differ a lot in their evaluations due to randomly selected first opinions.

Social contagion occurred a lot in our group dynamics. For instance, in one session, a pupil jumped in and claimed that Santa Claus was one of the characters in Picasso's *Saltimbancs*. He was fat, wore red clothes and had a bag. The adherence to the position was massive. Another pupil then suggested that it could be a clown. But the proponent of the Santa hypothesis argued that it could not be so because clowns do not work in the desert. The facilitator praised him for the good point and asked if Santa Claus usually is represented in the desert. In general, pupils are more willing to change the subject instead of openly changing their opinion. A minority of pupils pointed out more counterevidence. The character had no beard, and despite being red, his hat and clothes did not look like Santa's. The facilitator tried to help by inviting these pupils to voice their arguments more than once. However, the first impression was so strong that, even in the face of seemingly convincing counterarguments, the great majority of pupils in this class could not abandon their view. Proving that there is some arbitrariness in the dynamics, in other classes, the hypothesis of Santa either did not appear or appeared but was not particularly popular.[23]

We hypothesized that the prevalence of groupthink biases such as cascades might be connected to a shortcoming identified even among experienced practitioners of CI. They fail to take a knowledge-centered approach to intersubjectivity in which not only the multiple views are accounted for, but they do so under a joint goal of getting as close as possible to the truth of the matter.[24]

Reasoning was defined as thinking constrained by the inferential norms that one deems appropriate (1.1). Thus, we supposed that one way to prompt a knowledge-centered intersubjectivity was to make the relevant inferential norm collectively visible to the group. To do so, we would need some instruction concerning the methods of good reasoning and their justification as CT supports. However, we wanted to keep the success of habituation through a visual structure that is

23 The facilitator usually asks who wants to voice their opinions first. This design selects the more extroverted students. These students are likely to be more popular as well. Both factors increase the likelihood of cascades. See 34.

24 See intersubjectivity oriented towards knowledge in 17.

accessible to all participants of the community of inquiry. This hypothesis motivated our design of a routine to train children to make inferences to the best explanation.

Let us sum up the information that served as our guiding principles:

- Making the structure of thinking visible works, but biases are pervasive.
- The community approach helps diminish the egocentric perspective, but it brings its own set of social biases.
- A structured instruction of procedural techniques for critical thinking was missing and might help.

Before we proceed, we want to mention that we anticipate that defenders of each of the approaches above will point out that the shortcomings may come from our failure to follow the recommendations of each claim or the apparent time limitations of the application. Our response is a pragmatic one. Our application is too restricted to serve as evidence against any of their claims. However, we do believe that our experience identified important problems and that our reaction to them indicates promising solutions that might be useful for other contexts.[25] Reasoning is complex but has a determinable complexity. Educating for good reasoning will be unavoidably complex but should be so in a determinate way. Combining virtues education, community of inquiry and critical thinking offer a way to structure such a determinate complexity.

2 The Prototype

2.1 The natural outset

Familiarity and fallibility are useful pedagogical criteria to select a starting point for any teaching endeavor. We treat them as complementary conditions. Familiarity gives the educators something to build upon, while fallibility assures the need for learning. While reflecting on which specific technique we wanted to incorporate into a thinking routine, it became clear that the inference to the best explanation is both familiar and fallible. Undefined inference occurs almost naturally in human cognition. In our experience, children usually jumped to different conclusions based on unconscious inferences concerning some elements of the paintings. The process came so naturally to them that they tend not to perceive it.[26] This immediacy is part of the problem. Without awareness, children cannot properly evaluate their reasoning.[27] In our application, we noticed that many biases affected the children's assessment of the reasons they provided for their interpretations. Thus, we hypothesized that the inference to the best explanation could provide an effective way to combat these shortcomings.[28]

In line with habituation as defined above (1.2), our routine tries to mimic the occasions in which questions demanding an inference to the best explanation arise.

25 Neither are we claiming that there were no practices in CI, VE and CT that meet the criteria for what we are calling strong pluralism.

26 See 36.

27 Thus, even if they became perfect dispositional believers as in a VE utopia, it would not be enough.

28 Strictly speaking, the inference to the best explanation is not a formal method. Some philosophers even characterize it as fallacious, but its usefulness for scientific investigation is undeniable. Hence, the necessity of being self-aware of its potential shortcomings.

Instead of starting with formal instruction and then presenting the routine, we kept the initial steps of 'see, think and wonder'.

The first step was 'see'. The students (in groups of ~10) should carefully observe a painting. However, we included another feature. The children were asked to write their thoughts before the public discussion. This step should diminish the social contagion effects on their opinions. As research in decision-making confirms most of the benefits of having diverse perspectives in a group occur when individual opinions are truly independent [35].

After writing their thoughts, public discussion follows as usual. Children present their views, and the facilitator constantly asks the magic 'what makes you say that' questions. As previously identified, interpretative issues arise very early on these occasions. In Kirchner's *View of Basel and the Rhein,* some saw a church while others saw a castle and even a rocket! The facilitator takes the opportunity to choose the most promising polemical element in a painting and make the different views salient to the whole class. At this point, making the thinking visible begins. A table is made on the blackboard with a column for each hypothesis. The children also receive a copy of the table on a piece of paper which they need to fill in with all their hypotheses (see appendix 1). The observation, notes and the visible structure provided by the table set the stage for the formal instruction.

2.2 The formal instruction

We added a step of formal instruction to our routine to answer the CT requirement. Furthermore, instead of simplifying the vocabulary by avoiding jargon, we opted to try out and see if the children would learn to use words such as hypothesis and evidence.[29] They immediately nicknamed them the 'hard words', but they were keen to use them.[30]

Hypothesis

To explain what is a hypothesis, we offer near synonyms such as supposition, opinion, and others. However, the most useful pedagogical strategy consists in emphasizing the structure in which hypotheses present themselves. In this case, we say that a hypothesis is what happens in an 'I think that…' phrase.

The facilitator takes the opportunity to make it evident that all the interpretations on the table suit the model.

Look at the table. We can fit the 'I think that' phrase in each column. For instance, I think that… it is a rocket. Hence, these are our working hypotheses.

29 We also have a step of instruction on 'inference', but we did not include it here.

30 It is hard to know to what extent the children understand the concepts and their use as we present them. They clearly did not achieve a full-fledged use competence even after one year of activities. However, in our experience, the repetition that is a feature of a routine approach was key for the assimilation and progress. We repeated not only the routines but also the theoretic explanations. In addition to that, the in-person facilitator would spend the day with the children and use quotidian events as opportunities to point out occasions for using what they learned in the routines (see an example in 'Meta-cognition' below).

Evidence and Explanation

When faced with the question about the polemical element, the children tend to be very excited to defend their views, so much so that they express it by screaming their answers. Here is the opportunity for the facilitator to calm down the atmosphere and show that the proper way to defend a view is by giving reasons that support it.[31]

At this point, the moderator makes evident the full structure of a hypothesis and its reasons. We use the same strategy of offering a fixed structure with the content to be filled.

The reasons are what come after a 'because'. Thus, we have

I think that… [and we enter a hypothesis] because... [and here we provide a reason]

The tricky part is to differentiate the evidence that is accessible to all and the cognitive work leading to an explanation. Each pupil must realize that each explanation is an individual cognitive process. Thus, the others will only be able to grasp an explanation if it is made explicit.

Usually, giving a reason is giving evidence that makes clear to others what we are thinking. The evidence is the same for all and the 'because' is our reasoning. Take the explanation 'I think it is a rocket because it has a red triangular top.' Everybody can see the red triangular top, but not everyone will agree that this proves that it is a rocket. That is why we must explain the evidence, and the explanation must be convincing. Think about another explanation for the same evidence. For instance, 'the red triangular top is a roof'. Which one do you think is more convincing?[32]

The facilitator should provide several examples based on what the children previously expressed. Then the children are also invited to offer more evidence and explanations. While they do so, the cells of the table below the hypotheses are filled.

There is the possibility of inserting a rule of gesture to fix the structure. For instance, the pupil should raise their finger in connection to the 'I think that… (hypothesis)' and then turn it 90 degrees to make a point in connection to the 'because... (evidence)' as in **Fig.1**.

31 The facilitator must always navigate through the children's excess of excitement or lack of interest throughout these routines. Lack of interest tends to appear in the more theoretic explanations. Excessive excitement usually occurs when children want to defend their views. Hence, asking for children's opinions can make the environment livelier, including during theoretic explanations.

32 In the future, we may experiment with more refined divisions such as Toulmin's [37]: claim, data, warrants, backing, rebuttals and qualifiers. But, at first, it seemed too complex to be useful for the pupils.

***Fig. SEQ Figure * ARABIC1**: Gestures help fixing the structure.*

I think that… *because...*

Counter-evidence

Once evidences for the competing hypotheses are supplied, it is time for the facilitator to point out another way of arguing for a view.

We can also give counter-evidence that weakens a competing hypothesis.

The children do it naturally, so some examples tend to be available. Again, the facilitator only has to make it salient and structured. For instance,

It cannot be a rocket because it is in the middle of a city.

Then the children are invited to find counterevidence for each of the reasons they provided. The third row in the table is filled. The pupils now have a structured visualization that maps all the hypotheses and the reasons for and against them. This sort of argument mapping is in line with some trends in critical thinking [38]. Apart from that, up until here, there is nothing new in comparison to CI. But the main problem we wanted to address concerns assessing the arguments.

Assessment

To offer the pupils a procedural approach to assessing different arguments, we built upon recommendations from the research on judgment-making. We focused on three guidelines from cognitive psychology. (1) Our final judgments are more on point when the main question is divided into simpler ones. (2) We are better at relative judgments based on comparing options rather than absolute ones. (3) The adoption of a common scale helps to provide mutually comparable partial answers [28].

In our table, the choice of a hypothesis stands for the general question. The explanatory evidence break it up into more specific questions. Since these are divided into “for and against”, we are in a good position to make relative evaluations. To supply the missing feature, a scale, we present the pupils with a three-leveled system going from three stars for compelling reasons to one star for weak ones.[33]

33 We used a scale of three for simplicity. More trained groups might experiment with a 5 or 7 scale. More than seven is not recommended [39].

During the presentation, the facilitator points out the comparative nature of the assessment.

Since we have pros and cons, we need to judge how many stars we give to a reason in relation to one another. For instance, before giving the stars for the reason 'rockets have red and triangular tops', we have to compare it with the reason 'red triangular top is a roof and rockets do not have a roof'. The stronger should receive more stars. Other reasons can come to help. Thus, 'rockets do not stay in the middle of the city' might enter into the comparison and help us decide to put more stars for one or another reason.

The children should individually perform the assessment on their own tables first. We took as positive evidence that, at this point, most of the children have already changed their minds, and the less plausible hypotheses are tacitly abandoned.[34] However, more than one contender usually remains.[35] This is a positive feature since it mimics most of the situations that the children will face in their lives. For the educators, conflict brings the opportunity to make reasoning collective.

2.3 The Collective Assessment

Let us start with a quick overview. After a natural start for the routine, we had procedural instruction as recommended by CI. Then we made the structure of the elements and evaluation of inference visible in the table, following the Project Zero and the argument mapping approaches. According to the recommendations of cognitive psychology, we used comparisons to calibrate each pupil's degree of certainty concerning each explanation. Hopefully, this will also train their intuitions, as in VE. Finally, the stage is ready for approaching the inference to the best explanation from the perspective of collective reasoning. We predicted that this step and the previous one will decrease the groupthink biases.

The common scale of stars comes to hand in open discussions. The facilitators tell that the first shared goal is to try to agree on how many stars each reason deserves. After reading one of the reasons, we ask one of its defenders to suggest how many stars it deserves and argue for their level of certainty. Thus, we avoid extroverts always coming first. In the sequence, other pupils can express their agreement, doubts and objections. The collective enterprise is emphasized by contrasting the two phases.

When thinking about the evidence in favor of and against a hypothesis, each one of us defended our views. This is important. People who are for some hypothesis will be good at providing reasons for it, while people who are against it will be better at spotting the weak points.

The strongest version of each hypothesis is created by explaining the evidence (the things that all can see in the painting) and its interpretation. We also did that. Now, we have to select the best explanation. From this point on, it is no longer about trying

34 See 36. Children are good at evaluating the certainty of deduction. In particular after they become aware of their inferences.

35 There seems to be a tendency that more abstract and less precise paintings lead to more competing hypotheses.

to defend individual opinions anymore. All the information belongs to all of us as a group and, collectively, we must try to decide which one of them provides the best general interpretation of the issue, that is to say, which offers the best explanation.

A series of comparative discussions about the pros and the cons follow. At each step, we try to agree on how many stars should be given to each pro and con in the open collective table. In the end, we do a collective general appraisal of the stars in the whole table to estimate which hypothesis is the best. This sums up our proposal to make intersubjectivity visible.[36]

The development of this final design of the routine was the result of ongoing applications and adaptations. Thus, based on the need, we are already planning other applications to gather the necessary data for a steadier evaluation. However, the first impressions were promising. During the second semester, we noticed that the awareness of the pupils' own reasoning continued to increase. Also, exaggerated fancifulness in an association of ideas was spotted when making the evaluation and usually received just one star. The use of the stars applied to each reason provided a way for the pupils to convince themselves to abandon their first anchored general hypotheses. Much to the surprise of the pupils, when they added the results of their divided evaluations they recognized that the final result frequently ended up being different from their first general evaluation. Finally, the level of agreement as a group at the end of the sessions was higher than before. More importantly, only the more plausible hypotheses tend to survive the process. Thus, it seems that the power of arbitrary criteria was reduced, including the effect of a first voiced opinion.

Meta-cognitions

We hope that the open table filled individually but discussed and assessed collectively works as a map for collective metacognition. To emphasize the meta-level, after the routine, the facilitator openly reviews what the class has made [40]. While doing so, the role of each part of the table should also become clear. Thus, the children are visually exposed to the metacognitive awareness of the technique.

It all started when we saw that the same element in a painting led to different conclusions. After hearing the others, we realized that there were several plausible hypotheses. To try to find the best one, we filled the first row with all the hypotheses. Then we had to investigate what were the sources of our hypotheses. We also had to explain them to the others to show how compelling they were. We also saw that some of us saw that some explanations did not seem to work that well. These problems are the cons or the counterarguments with which we filled in this column. After filling it, we used the stars to evaluate each piece of information. But we always did so in comparison with the counterargument and other explanations. After some discussion, we arrived at a very confident answer.

It is also important to show that the same process can be used in their quotidian. During the semester, we had an event that provided a convenient occasion to do so.[37]

36 If there were close cases, we were OK with letting the matters open. Also, if some of them want to stick to any other hypotheses, they can do so.

37 The use of this and other situations (such as noticing how the movies that the children saw influenced their interpretations, see 'First Results' above) was only possible because Débora, one of the facilitators, interacted

We can use this table and this scale of values in all sorts of situations in our everyday life. For instance, last week the 4th grade built a city in our garden. The following day, the 3rd grade entered the play and added some buildings! But after the weekend, the city was destroyed. Some of you said that the 3rd grade was responsible for that. You can resort to the table of the best explanation to solve the matter.[38]

The final review provides an opportunity to praise the children for their good behavior. In this context, the facilitator emphasizes the necessary intellectual virtues for the activity, open-mindedness and intellectual humility. Again, there is the occasion for nurturing self-awareness

Intellectual humility is necessary for us to admit that our initial positions were not strong as we supposed they were before arguing for them. Without open-mindedness, we cannot seriously engage with the other hypotheses, reasons and counter-arguments.

Finally, it is also time to increase self-awareness about the importance of collective enterprise. At this point, it is worth emphasizing how the group arrived at a hypothesis stronger than the sum of the individual opinions and that this is only possible because of using the right technique, with the right disposition, and in the ideal environment.

Conclusion

To conclude the report, we want to reinforce that, after six months of application, the results, although remarkable to us, must be considered carefully. The overall quality of the arguments and even children's attitudes improved. They also became more and more at ease with identifying the origins of their conclusions. Of course, these positive results probably have to do with the overall experience, including the first six months of routines. More specifically, in the second phase, the children started using new vocabulary such as hypothesis and evidence, including in other contexts. The system of evaluation of each reason in comparison seems to be the source of much improvement since the good results became evident even before the collective step. However, there is no denying that the collective discussion increases the likelihood of the prevalence of the more plausible hypotheses.

The six months of application was a prototyping period. Now that we have arrived at a final set-up, we plan to apply this routine in different schools and contexts to gather more feedback. We also hope that this publication will reach an audience that may be willing to apply some versions of it. More generally, we are also interested to see if the triadic approach to the routines with well-defined steps to combine reasoning, a reasonable disposition, and reasonableness can live up to our expectations. Following that, we also expect to develop other routines that deal with different specific techniques of good reasoning.

with the children in several of their activities daily. We also informally discussed the possible effects of the routines with the teachers of the project.

38 The children concluded that the 3rd grade would not destroy something they helped create.

Final quote: Reasoning is complex but has a determinable complexity. Educating for good reasoning will be unavoidably complex but should be so in a determinate way. Combining virtues education, community of inquiry and critical thinking offer a way to structure such a determinate complexity.

References

[1] Breidenstein G, Hirschauer S, Kalthoff H, Nieswand, B (2013) Ethnografie - Die Praxis der Feldforschung. UTB, Stuttgart

[2] Splitter L J, Sharp A M (1996) La otra educación Filosofía para Niños y la comunidad de indagación. Manantial, Buenos Aires

[3] Baehr, J (2015) Cultivating Good Minds: A Philosophical & Practical Guide to Educating for Intellectual Virtues. Intellectual Virtues, Maryland

[4] Roberts R, Wood W J (2007), Intellectual Virtues. OUP, Oxford

[5] Tishman S, Jay E, Perkins D E (1993) Teaching thinking dispositions: From transmission to enculturation. Theory Into Practice 32(3):147-153

[6] Bailin S, Siegel, H (2003) Critical Thinking. In: Blake N et al (eds) The Blackwell Guide to the Philosophy of Education, p 181-193

[7] Kotzee B, Carter J A, Siegel H (2019) Educating for intellectual virtue: a critique from action guidance. Episteme 18(2):177-199

[8] Moshman D (2004) From inference to reasoning: The construction of rationality. Thinking & Reasoning 10(2):221-239

[9] Siegel H (2017) Critical Thinking and the Intellectual Virtues, Intellectual Virtues and Education. In: Baehr J (ed) Essays in Applied Virtue Epistemology. Routledge, London p 95-112

[10] Baehr, J (2011) The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology, OUP, Oxford

[11] Baehr, J (2013) Educating for Intellectual Virtues: From Theory to Practice. Journal of Philosophy of Education 47(2): 248-262

[12] Bevan, R (2009) Expanding Rationality: The Relation between Epistemic Virtue and Critical Thinking. Educational Theory 59(2): 167-179

[13] Battaly, H (2016) Responsibilist Virtues in Reliabilist Classrooms In: Baehr J (ed) Intellectual Virtues and Education: Essays in Applied Virtue Epistemology. London, Routledge

[14] Ritchhart R, Church M, Morrison K (2011) Making Thinking Visible. Jossey-Bass, San Francisco

[15] Reznitskaya A, Glina M, Carolan B, Michaud O, Rogers J, Sequeira, L (2012) Examining transfer effects from dialogic discussions to new tasks and contexts. Contemporary Educational Psychology 37 (4): 289-306

[16] Poce, A (2021) Virtual museum experience for critical thinking development: First results from the National Gallery of Art (MOOC, US). Journal of Educational, Cultural and Psychological Studies 24: 67-83

[17] Daniel M-F, Splitter L, Slade C, Lafortune L, Pallascio R, Mongeau P (2004) Dialogical critical thinking: Elements of definitions emerging in the analysis of transcripts from pupils aged 10 to 12 years. Australian Journal of Education 48(3): 295-313

[18] Lipman M (2003) Thinking in education. CUP, Cambridge

[19] Lipman M, Sharp A, Oskanyan F S (1980) Philosophy in the Classroom. Temple University Press, Philadelphia

[20] Battaly, H (2006) Teaching Intellectual Virtues. Teaching Philosophy 29(3): 191-222

[21] Mizrahi, M (2013) Why arguments from expert opinion are weak arguments. Informal Logic 33: 57–79

[22] Hookway, C (2000) Regulating Inquiry: Virtue, Doubt, and Sentiment. In: Axtell G (ed) Knowledge, Belief, and Character: Readings in Virtue Epistemology. Rowman & Littlefield, Lanham p 149-162

[23] Zagzebski, L (2017) Exemplarist Moral Theory. OUP, Oxford

[24] Annas (2011) Intelligent Virtue. OUP, Oxford

[25] Novakowski D, Mishra S (2018) Biases. In: Shackelford T, Weekes-Shackelford V (eds) Encyclopedia of Evolutionary Psychological Science. Springer, Cham. https://doi.org/10.1007/978-3-319-16999-6_627-1

[26] Carter J A, Pritchard D (2016) Cognitive Bias, Scepticism and Understanding. In: Grimm S R, Baumberger C, Ammon S (eds) Explaining Understanding. Routledge, London p 272-292

[27] Mărăşoiu A I (2020) Intellectual Virtues and Biased Understanding. Journal of Philosophical Research 45: 97-113

[28] Kahneman D, Sibony O, Sunstein C (2021) Noise: A Flaw in Human Judgment. Little, Brown Spark: New York

[29] Kuhn D, Shaw V, Felton M (1997) Effects of dyadic interactions on reasoning. Cognition and Instruction 15(3): 287-315

[30] Osborne, J, Erduran, S, Simon, S. (2004). Enhancing the quality of argumentation in school science. Journal of Research in Science Teaching 41(10): 994-1020

[31] Bang D, Frith, C D (2017) Making better decisions in groups. Royal Society open science 4(8): 170-193

[32] Salganik M, Watts D (2008) Leading the Herd Astray: An Experimental Study of Self-fulfilling Prophecies in an Artificial Cultural Market. Social Psychological Quarterly 71: 338-355

[33] Brem S, Rips L (2000) Explanation and evidence in informal argument. Cognitive Science 24(4): 573-605

[34] Robalino J D, Macy M (2018) Peer effects on adolescent smoking: Are popular teens more influential PLoS ONE 13(7): 1-12

[35] Davis-Stober C P, Budescu D V, Broomell S B, Dana J (2015) The Composition of Optimally Wise Crowds. Decision Analysis 12(3): 130-143

[36] Pillow B H, Hill V, Boyce A, Stein, C (2000) Understanding inference as a source of knowledge: Children's ability to evaluate the certainty of deduction, perception, and guessing. Developmental Psychology 36: 169-179

[37] Toulmin S E (1958) The Uses of Argument. CUP, Cambridge

[38] Gelder T V (2005) Teaching Critical Thinking: Some Lessons From Cognitive Science, College Teaching, 53(1): 41-48

[39] Miller G A (1956) The magical number seven plus or minus two: some limits on our capacity for processing information. Psychological review 63(2): 81-97

[40] Berkowitz M W, Bier, M C (2007) What Works in Character Education. Journal of Research in Character Education 5(1): 29-48

Appendix 1

Table: Inference for the Best Explanation

Hypotheses	1.	2.	3.	4.
Pro				
	-			
stars				
	-			
stars				
	-			
stars				
Con				
	-			
stars				
	-			
stars				
	-			
stars				

Stars scale: * weak; ** medium; *** strong