# Deductive Reasoning and Learning: a Cross-Curricular Study. Research Report

Research report number 2(4)/2014; published: May 29, 2014. This work is a part of the project titled *Erotetic logic in the modeling of ultimate and distributed question processing. Theoretical foundations and applications* supported by funds of the National Science Centre, Poland, (DEC-2012/04/A/HS1/00715).

Mariusz Urbański(1)
Katarzyna Paluszkiewicz(1)
Joanna Urbańska(2)

(1) *Department of Logic and Cognitive Science*
*Institute of Psychology, Adam Mickiewicz University*

(2) *Department of Foundations of Psychological Research*
*Institute of Psychology, Adam Mickiewicz University*

2014
Poznań, Poland
*www.kognitywistyka.amu.edu.pl/intquestpro*

# Deductive Reasoning and Learning: a Cross-Curricular Study

Mariusz Urbański, Katarzyna Paluszkiewicz, and Joanna Urbańska

Institute of Psychology, Adam Mickiewicz University, Poznań, Poland
{Mariusz.Urbanski, Katarzyna.Paluszkiewicz, Joanna.Urbanska}
@amu.edu.pl

**Abstract.** We report a research on correlations between level of fluid intelligence and fluencies in two kinds of deductions, simple (syllogistic reasoning) and difficult ones (erotetic reasoning), and on the impact of learning logic on fluencies in deductive reasoning. We observed that although participants of higher ability (with respect to fluid intelligence) performed better than participants of lower ability in both kinds of deductive tasks, those who undergone extensive training in formal logic obtained significantly higher results than the rest of this group in tasks involving difficult deductions. We conclude on this basis that fluency in difficult deductions, while related to fluid intelligence, is related also to subjects' educational experience and that this does not hold in case of simple deductions.

**Keywords:** fluid intelligence, deductive reasoning, erotetic reasoning, learning logic.

## 1 Introduction

In classifications of reasoning based on logical entailment one usually distinguishes at least deductive and ampliative (or non-deductive) reasoning. The first is understood as referring to all reasoning that entails the truth of the conclusion given the truth of the premises, the second as all reasoning that does not entail the truth of the conclusion given the premises. These two kinds of reasoning serve different purposes and differ with respect to their logical structure. We are not going to enter the discussions on whether deductive reasoning involves unique psychological processes or not (Ricco and Overton [2011]) and on possibly limited role of deductive processes in ordinary cognition (Oaksford and Chatter [2007]). We take for granted the notion of deduction so construed and seek for correlations between deductions of different complexities and fluid intelligence, as well as for the impact of learning formal logic on fluency in deductive reasoning.

In line with two-components models of intelligence (Cattell [1971], Carroll [1993], Horn and Noll [1997]), it is generally assumed that fluencies in both deductive and ampliative reasoning are rooted in fluid intelligence (Gf) – which is the ability to reason form concepts, and solve problems that often involve unfamiliar information or procedures (Horn and Noll [1997]). The defining common feature of Gf reasoning tasks is that they require an individual to organize information, whether novel or familiar, and to identify patterns or rules for grouping or classifying information (Blair [2010]). Crystallized intelligence (Gc), in turn, comprises the skills and knowledge acquired through experience, education and acculturation (Cattell [1971]).

The precise location of deductive reasoning ability on the map of intelligence is not uniquely determined. On the one hand, fluid intelligence is conceptualized as being ordered along a continuum from elementary awareness to immediate memory to working memory to inductive reasoning to deductive reasoning (Horn and Noll [1997]). On the other hand, Carroll's [1993] metaanalysis of reasoning studies reveals that deductive reasoning factor loads fluid as well as crystallized intelligence, and the mixture of both (cf. Wilhelm, Engle [2005, p. 382–383]).

Nevertheless, according to Cattell-Horn-Carroll (CHC) model of intelligence (McGrew [1997], Schneider and McGrew [2012]) – currently probably the most influential psychometric model of intelligence – deductive reasoning (or general sequential reasoning) is identified as one of the three reasoning abilities defining fluid intelligence. The other two abilities are induction and quantitative reasoning, although whether the last one should be subsumed under fluid intelligence is somewhat controversial (see Wilhelm, Engle [2005]).

Different perspective is offered by proponents of dual-process accounts of reasoning (Sloman [1996], Evans [2003], Stanovich [2011]), who argue that deductive inferences are to be attributted to Type 2 processing. It is ,,considered to be high effort, responsive to instructions, and capable of dealing with abstraction and novelty, often allowing at least high-ability participants with appropriate instructions to solve deductive problems", whereas Type 1 processing is ,,rapid and low effort, and may be based on associative processing and implicit knowledge" (Evans [2012 p. 16]). Stanovich [2008] suggested an additional distinction within Type 2 processing of algorithmic and reflective (or intentional) component. Ricco and Overton [2011] developed this distinction with respect to deductive reasoning in their Competence $\leftarrow - \rightarrow$ Procedural processing model. They argue that ,,increasing organizational complexity at the algorithmic level is paralleled by progressively more explicit conceptual knowledge about logical categories (metalogical knowledge) at the reflective level" (Ricco and Overton [2011 p. 139]). Thus a question arises: what is the impact of Gf on both components in the case od deductive reasoning?

Certainly, in the research on reasoning there is some confusion, terminological as well as conceptual. On the one hand, what is called 'induction' in the psychological literature is in fact either abductive reasoning (although mistaken applications of this term also can be found; see Singmann and Klauer [2011, p. 251] for an example), whether explanatory or not (see e. g. Magnani [2009]), or ampliative reasoning. On the other hand, classification of reasoning into deductive and ampliative ones, based on logical entailment, is somewhat orthogonal to existing typologies of reasoning abilities, as the one proposed in the CHC model (Ajdukiewicz [1974] offers a good example of such classification, although he defines slightly more general concepts of subjectively certain inference vs subjectively uncertain inference). The reason is that from a logical point of view some complex forms of reasoning involve both deductive and ampliative reasoning as their components (explanatory abduction being a good example of such complex reasoning; see Gabbay and Woods [2005]). One may argue – quite aptly – that logical perspective does not offer a particularly good account on human reasoning (see Evans [2012] for insightful remarks on 'logicism' in the psychology of reasoning). Nevertheless, clear understanding of what is meant by 'deductive reasoning' and 'non-deductive reasoning' is crucial for establishing a robust theoretical framework for this kind of research. It hardly comes as a suprise to find out, for example, that fluency in abstract reasoning is correlated with insight problem solving ability (Paulewicz et al. [2007]) if one realizes that more often than not such problem solving tasks involve complex reasoning which subsume deductions as well.

Our assumption was that some deductions are easy (as in simple syllogistic reasoning in abstract settings) and some are more difficult (as in complex conditional reasoning). The term 'difficulty' here is to be understood qualitatively rather than quantitatively, although empirical justification for this distinction is solid (see table 1). Our hypothesis was that there is a strong correlation between Gf and fluency in easy deductions, whereas in case of more difficult deductions reasoners' experience should have a noticeable impact on their performance in such tasks. It is generally accepted that fluid intelligence increases until some point of the adult life span and then it starts to decline and that the effects of this decline may be partially compensated by development of crystallized intelligence (McArdle et al. [2002], Sternberg [2005]). Whether fluid intelligence can be improved by training or not remains an open problem, although more and more research suggest that training can improve fluid intelligence in both children (Klauer and Willmes [2002], Bergman Nutley et al. [2011]) and adults (Jaeggi [2008], see also Sternberg [2008] for an interesting commentary; it should be noted that the results of this study are highly controversial, e.g. Choi and Thompson [2012] conducted a non-replication of it). Thus we wanted also to find out what, if any, is the impact of learning on the level of Gf and fluencies in both kinds of deductions.

Tasks involving syllogistic inferences are a well-known tool in research on reasoning. Logical framework for difficult deductive tasks in this research was set up by Inferential Erotetic Logic (IEL; Wiśniewski [1995, 2013]), which allows for modeling erotetic reasoning, that is, reasoning involving questions either as conclusions or as both premises and conclusions. IEL-style logic of questions can be based on any logic which satisfies some simple syntactic and semantic conditions. Consequently, IEL allows for formalization of erotetic reasoning at different levels of complexity of specification of considered verbal representations. We are interested in inferences in which conclusion and one premise are questions and other premises – if there are other premises at all – are declarative sentences (these are so-called erotetic inferences of the second kind; see Wiśniewski [1995, p. 209]). Validity of such inferences is defined in terms of semantic relation of erotetic implication, which meets the following conditions:

1. transmission of truth/soundness into soundness: if the question-premise is sound (i.e., there exists a true direct answer to this question) and all the declarative premises (if any) are true, then the question-conclusion is sound as well;
2. cognitive usefulness: each answer to the question-conclusion is useful in answering the question-premise (each answer to question-conclusion narrows down the class of possible answers to question-premise), provided that all the declarative premises (if any) are true.

Erotetic inferences are good representations of some techniques of problem solving, either by reduction of an initial problem to a simpler one(s), or by identifying missing information which is needed in order to solve the initial problem (Urbański, Łupkowski [2010]). Consider a simple example (taken from Urbański, Łupkowski [2010, p. 68]). Suppose that our problem is expressed by the initial question:

($Q$) Who stole the tarts?

Suppose also that we managed to establish the following evidence:

($E_1$) It is one of the courtiers of the Queen of Hearts attending the afternoon tea-party who stole the tarts.

Thus the initial question together with the evidence erotetically implies the question:

($Q^*$) Which of the Queen of Hearts' courtiers attended the afternoon tea-party?

It is intuitively justified to ask for the list of courtiers – participants of the afternoon tea-party ($Q^*$) in order to solve the problem ($Q$), in view of the established evidence ($E_1$). This justification can be expressed in exact terms by fulfillment of both conditions of erotetic implications. First, if somebody really stole the tarts and if it is true, that the culprit is one of the courtiers of the Queen of Hearts attending the afternoon tea-party (that is, if $Q$ is sound and if $E_1$ is true), then some of the courtiers must have attended the party (that is, $Q^*$ is sound as well). Second, each non-empty list of courtiers – participants of the party narrows down the class of suspects, provided that it is really one of the courtiers of the Queen of Hearts attending the afternoon tea-party who stole the tarts (that is, each direct answer to $Q^*$ narrows down the class of possible answers to $Q$, in view of $E_1$).

If moreover we know that:

($E_2$) Queen of Hearts invites for a tea-party only these courtiers who made her laughing the previous day.

then $Q^*$ and $E_2$ erotetically imply the question:

($Q^{**}$) Which courtiers made the Queen of Hearts laughing the previous day?

It is easy to check that in this case both conditions of erotetic implication are fulfilled as well.

Such erotetic inferences clearly involve deductive reasoning (especially in view of the first condition imposed on erotetic implication). However, the presence of verbal representations different than the usual declaratives, and the presence of an additional condition of cognitive usefulness suggest that carrying out erotetic reasoning may be a more difficult task than carrying out simple syllogistic reasoning (Stenning and van Lambalgen [2008]).

## 2 Method

Our research were carried out between February and May, 2012. 105 subjects were recruited (M = 21.69, SD = 1.44, 88 women), students of Adam Mickiewicz University in Poznań (major Polish city and academic center) and of State School of Higher Professional Education in Gorzów (an important administrative center in western Poland), who volunteered to participate in this research. They represented the curricula of cognitive science, psychology and educational studies.

The choice of the curricula was determined by three factors:

1. The same scientific area affiliation (in the Polish system all three are classified as social science curricula and their programmes are set up according to similar general principles).
2. The extent of formal logic courses (cognitive science curriculum offers 6 compulsory courses in logic, ranging from an introductory one to advanced proof methods and epistemic logic courses; psychology and educational studies curricula offer introductory course only).
3. Selectivity of admission procedures, based on A-level examination results (psychology and cognitive science admission procedures are highly selective).

As a result, three groups were distinguished:

A – extensive logic, selective admission (cognitive science; N = 27);
B – basic logic, selective admission (psychology; N = 40);
C – basic logic, non-selective admission (educational studies; N = 38).

Apparently, there is no social science curriculum of 'extensive logic, non-selective admission'-type in Poland.

None of participants received any extra-curricular training in logic. It should be noted that participants from groups A and B received, as a part of their curricula, a training in statistics (the same in case of both groups). Participants from group A received also some training in programming languages.

Motivation for such a choice of subjects was that participants from groups A and B, having achieved similar (and high) results in A-level examinations, were expected to exhibit similar levels of fluid intelligence, while in case of group C this level should be significantly lower. The same was expected in case of simple deductive reasoning. In the case of more complex deductions, involved in erotetic reasoning, participants from group A were expected to perform significantly better than participants from two other groups.

To each participant we administered three tests in one 90 minute session (in groups of ca. 20 persons): Raven's Advanced Progressive Matrices test (APM, Raven [2003]), Deductive Reasoning Test (DR) and Erotetic Reasoning Test (ER). Both reasoning tests were designed by the authors and were carried out in Polish. Statistical analysis was performed using IBM SPSS Statistics, version 21.

### 2.1 Raven's Advanced Progressive Matrices test

Raven's Advanced Progressive Matrices test (APM) contains 48 items, presented as one set of 12 (training set, no time limit), and another of 36 (time limit 30 min). In each item, the subject is asked to identify the missing element that completes a pattern. Each item is a $3 \times 3$ matrix of figural stimuli that are organized according to latent rules. A subject's aim is to discover proper rules and to choose one correct answer from eight possible ones. Items become increasingly difficult as progress is made through each set.

Raven's Progressive Matrices results are generally thought of as a good measure of reasoning ability component of general intelligence, especially its fluid factor (Blair [2010]). We used the advanced version in this research (Cronbach's $\alpha = .83$); if untimed, this test is designed to differentiate between people at the high end of intellectual ability. When administered under timed conditions, the APM can also be used to assess intellectual efficiency (Raven [2003]).

### 2.2 Deductive Reasoning Test

Deductive Reasoning Test (DR) contains 6 items (time limit 20 min). Each item consists of two declarative pieces of information and the task is to answer a question on the basis of the information given. These are in fact syllogistic premises and a question concerning which conclusion is justified by the premises. The chosen format allows for not using the notions of premise and conclusion (see Appendix 1 for English translation of DR tasks).

Three items are of categorical character and three of relational character, all set in abstract context using pseudowords for object categories and real names for relations. Some premises are affirmative and some are negative, but all questions concerning possible conclusions are based on affirmative sentences only. The semantic task of interpreting the meaning of negative answer to a question based on a negative sentence seems to overly complicate test items.

There is only one correct answer in each item. The criterion of correctness is the recognition of logical entailment.

Cronbach's $\alpha = .64$ occured to be just satisfactory for DR in this research. This is understandable, as the test contains only six items. Removal of any item reduces $\alpha$ coefficient. Guttman's $\lambda 4$ for DR was .71 while $\lambda 2$ was .65. DR turned out to be relatively easy for the subjects ($\bar{T}$ = 81,28%; see table 1).

### 2.3 Erotetic Reasoning Test

Erotetic Reasoning Test (ER) contains 3 items (time limit 30 min). Each item consists of a detective-like story in which the initial problem and evidence gained are indicated. The task is to pick a question (one out of four), each answer to which will lead to some solution to the initial problem. The subjects are asked to justify their choices (see Appendix 2 for English translation of an exemplary ER task).

All three stories describe some investigation and they invoke search for a solution to an initial problem by means of posing further (auxiliary) questions. The stories are set up in such a way that the impact of previous content-related experience of the subjects on the choice of solution is minimized. All the relevant information is explicitly listed and the subjects are asked to solve each task (i.e., to pick a correct question) solely on the basis of what is given.

There is only one correct answer in each item. However, the criterion of correctness is more complicated than in the case of DR. What matters is not only correct choice of the question-solution but also a proper justification of the choice, based on two conditions of validity of erotetic implication: transmission of truth/soundness into soundness and cognitive usefulness. Thus overall correctness of an answer in each ER task was based on the presence of these three elements. Reliability of the ER test turned out to be acceptable ($\alpha = .78$). ER occured to be generally substantially more difficult for the subjects than DR ($\bar{T}$ = 61,08%; see table 1), with the exception of group A (in this case difficulties of ER and DR were comparable).

## 3 Results

Tests scores statistics (table 1) show that APM and DR were more difficult for group C than for groups A and B. However, while results of groups A and B in these tests were similar, the results of group A were less scattered. ER turned out to be very easy for group A (and substantially easier than APM), but very hard for group C (and much more difficult that the remaining two tests).

A one-way multivariate analysis of variance (MANOVA) was conducted to test group differences with respect to three response variables (APM, DR and ER test scores). Following suggestions by Tabachnick and Fidell [2007] we decided that this is still a more adequate choice than carrying out a series of ANOVAs or some non-parametric test. Preliminary assumptions checking revealed that the data were normally distributed only in case of APM, in groups as well as in the whole body of subjects (Shapiro-Wilk test, $p > .05$). There were some univariate outliers in the data, as assessed by inspection of a boxplot: four in case of group A and DR, four in case of group A and ER, also four in case of group C and ER, and one in case of group B and APM. We decided not to remove them, as this

**Table 1.** Tests scores statistics

| Tests and groups | Min | Max | Mean | SD | 95% CI | Difficulty Index $\bar{T}$ |
|---|---|---|---|---|---|---|
| APM | 7 | 36 | 22.19 | 5.63 | | 61,46% |
| group A | 21 | 32 | 25.81 | 2.73 | [24.73, 26.90] | 71,60% |
| group B | 14 | 36 | 24.25 | 4.58 | [22.79, 25.71] | 67,35% |
| group C | 7 | 26 | 17.45 | 4.93 | [15.83, 19.07] | 48,48% |
| DR | 0 | 6 | 4.93 | 1.35 | | 81,28% |
| group A | 5 | 6 | 5.81 | .40 | [5.66, 5.97] | 96,92% |
| group B | 4 | 6 | 5.53 | .70 | [5.31, 5.74] | 91,67% |
| group C | 0 | 6 | 3.68 | 1.42 | [3.22, 4.15] | 61,46% |
| ER | 0 | 3 | 1.73 | 1.09 | | 61,08% |
| group A | 2 | 3 | 2.78 | .42 | [2.61, 2.95] | 92,60% |
| group B | 0 | 3 | 1.98 | 1.03 | [1.65, 2.30] | 65,83% |
| group C | 0 | 3 | .74 | 1.23 | [.39, 1.08] | 24,80% |

might significantly affect the results of the analysis. There were no multivariate outliers in the data, as assessed by Mahalanobis distance ($p < .001$). There was no multicollinearity (Pearson correlation: APMxDR $r = .580$, $p < .001$; APMxER $r = .692$, $p < .001$; DRxER $r = .701$, $p < .001$). The relationships were monotonic, as assessed by scatterplot. The assumption of homogeneity of variance-covariance matrices was violated, as assessed by Box's test of equality of covariance matrices (p < .001). The assumption of homogeneity of variances was violated for each DR, ER and APM score (Levene's Test, $p < .05$). There was a statistically significant difference between the groups on the combined response variables, F(6, 202) = 16.613, $p < .001$; Pillai's Trace = .661; partial $\eta^2 = .330$.

Follow-up univariate ANOVAs showed that scores in all three tests were statistically significantly different between participants from all the three groups, using a Bonferroni adjusted $\alpha$ level of .017 (see table 2).

**Table 2.** Group differences (one-way ANOVAs)

| Test | df | $F$ | $p$ | $\omega^2$ |
|---|---|---|---|---|
| APM | 2 | 36.809 | .001 | .405 |
| | 102 | | | |
| DR | 2 | 49.948 | .001 | .482 |
| | 102 | | | |
| ER | 2 | 40.750 | .001 | .431 |
| | 102 | | | |

Games-Howell post-hoc analysis revealed that in both APM and DR there were significant differences between groups A and C as well as between groups

B and C. In case of ER there were significant differences between all the three groups (table 3).

**Table 3.** Games-Howell test results

| Test and groups | Mean increase | 95% CI | $p$ |
|---|---|---|---|
| APM | | | |
| A x B | 1.56 | [-.58, 3.71] | .195 |
| A x C | 8.37 | [6.07, 10.67] | .001 |
| B x C | 6.80 | [4.22, 9.38] | .001 |
| DR | | | |
| A x B | .29 | [-.03, .61] | .079 |
| A x C | 2.13 | [1.54, 2.72] | .001 |
| B x C | 1.84 | [1.23, 2.45] | .001 |
| ER | | | |
| A x B | .80 | [.37, 1.24] | .001 |
| A x C | 2.04 | [1.58, 2.50] | .001 |
| B x C | 1.24 | [.67, 1.80] | .001 |

Discriminant analysis confirmed those results. It revealed two discriminant functions: the first explained 94,8% of the variance (canonical $R^2 = .77$), the second explained 5,2% of the variance (canonical $R^2 = .27$). In combination these discriminant functions significantly differentiated the groups, $\Lambda = .38$, $\chi^2(6) = 97.17$, $p < .001$. The second function significantly differentiated the groups as well, $\Lambda = .93$, $\chi^2(2) = 7.66$, $p = .02$. The correlations between outcomes and the discriminant functions revealed that ER results loaded onto both functions ($r = .73$ for the first function and $r = .68$ for the second), whereas DR and APM results loaded substantially onto the first function only (DR: $r = .83$ for the first function and $r = -.21$ for the second, APM: $r = .71$ for the first function and $r = -.03$ for the second). The discriminant function plot showed that the first function differentiated group C from groups A and B, and the second function discriminated group A from group B.

A Spearman's rank-order correlation was run to assess the relationship between tests results (table 4). Interestingly, while in the whole group results in all the three tests significantly correlated with each other, correlations in groups A, B and C followed different pattern. These results suggest that, although Gf and fluencies in both kinds of deductions are generally correlated, the scheme of these correlations is kind of complicated.

Correctness criterion in case of ER tasks is more complex than the one of DR tasks. In particular, explicit formulating of justification of the solution's choice may be seen as difficult, especially in case of less able subjects. Thus we carried out all the statistics again, with adjusted correctness criterion for ER tasks: just correct choice of a solution was taken into account (we shall refer to this adjusted version of the test results as to $ER_{ad}$). The effect just mentioned is confirmed by inspection of test scores statistics (table 5). It reveals that while justification

**Table 4.** Correlations $r_S$ between tests results (*$p < .05$, **$p < .01$)

| Tests | Overall | A | B | C |
|---|---|---|---|---|
| APM x DR | .61** | .22 | .24 | .23 |
| APM x ER | .67** | .08 | .42** | .59** |
| DR x ER | .74** | .43* | .55** | .41* |

requirement had some impact on results of groups A and B, in case of group C its impact was very substantial.

**Table 5.** Tests scores statistics – ER and $\mathrm{ER}_{ad}$

| Tests and groups | Min | Max | Mean | SD | 95% CI | Difficulty Index $\bar{\mathrm{T}}$ |
|---|---|---|---|---|---|---|
| ER | 0 | 3 | 1.73 | 1.09 | | |
| group A | 2 | 3 | 2.78 | .42 | [2.61, 2.95] | 92,60% |
| group B | 0 | 3 | 1.98 | 1.03 | [1.65, 2.30] | 65,83% |
| group C | 0 | 3 | .74 | 1.23 | [.39, 1.08] | 24,80% |
| $\mathrm{ER}_{ad}$ | 0 | 3 | 2.11 | 1.01 | | |
| group A | 2 | 3 | 2.85 | .36 | [2.71, 3.00] | 95,07% |
| group B | 0 | 3 | 2.15 | .92 | [1.86, 2.44] | 71,67% |
| group C | 0 | 3 | 1.55 | 1.08 | [1.20, 1.91] | 51,27% |

Again, there was a statistically significant difference between the groups on the combined response variables, F(6, 202) = 16.518, $p < .001$; Pillai's Trace = .658; partial $\eta^2 = .329$. One-way ANOVA showed that although scores in $\mathrm{ER}_{ad}$ were statistically significantly different between participants from all the three groups (F(2, 102) = 17,059, $p < .001$), effect size in this case was much lower ($\omega^2 = .0234$ as compared to .431 for ER; see table 2). Statistically significant differences, revealed by Games-Howell post-hoc tests, were similar as in the previous analysis, i.e. when overall correctness for ER was composed of correct choice and two-element justification.

## 4 Discussion

High correlation between APM and DR results in the whole body of subjects and lack of significant in-group correlations allow for a conclusion that although the measured abilities are generally related to each other, subjects in groups displayed substantial individual differences in their levels. This holds for APM results in case all the three groups and for DR results in case of groups B and C. DR scores statistics show that almost all participants from group A solved all tasks correctly (in fact only four of them did not: they were the DR univariate outliers we reported above).

There were moderate to high correlations between APM and ER results in groups B and C, while results of group A lacked such correlation. Also, results of group A in ER were significantly higher than results of group B and there was no such difference in case of DR and APM, while both group A and group B outperformed group C in DR, ER, and APM. These results show that in case of participants from group A their performance in ER was facilitated by some factor, which was not present in case of participants from groups B and C. In our opinion this suggests that although solving ER tasks is related to Gf, it is related also to subjects' experience: what we have observed is, most probably, the impact of learning formal logic (and related subjects) on fluency in solving certain type of problems. Obviously, our correlational study does not allow for drawing conclusions concerning causal links; what we can conclude is that, on the one hand, abilities to carry out simple and complex deductions are related to each other, as witnessed by moderate to high correlations between DR and ER results. On the other hand, deductions of different complexities call for different abilities to be manifested and fluency in more difficult deductions is related not just to the subjects' fluid intelligence, but to the subjects' experience as well.

An obvious limitation of this conclusion is that in order to confirm it substantially a study with robust measures of crystallized intelligence is needed, in order to establish a pattern of correlations between the level of Gc and fluency in erotetic reasoning. One posibility is to use the Verbal Comprehension Index subscale of WAIS-IV (Wechsler [2008]), which has been shown to measure Gc (Benson et al. [2010]). Unfortunately, WAIS-IV is not yet available in Polish.

Another possible explanation of the results is that what we have observed is an impact of learning indeed, but just task-specific one. One possibility is, that task specificity is related to deductive character of the reasoning involved; but in such case there should be also a significant difference between groups A and B in solving DR. The other possibility is, that the subjects from group A learned to carry out only specific kind of difficult deductions. This hypothesis can also be tested by administering some general intelligence test (like WAIS-IV) to the subjects. However, while all the subjects received (as parts of their curricula) basic training in syllogistic reasoning and in Classical Propositional Calculus, they received no training whatsoever in erotetic reasoning. Thus the learning process in case of group A resulted not in mastering certain kind of tasks, but in improving related problem-solving abilities; and this is essentially another wording of our first conclusion.

An important yet often overlooked distinction in research on reasoning is that between reasoning *to* vs reasoning *from* an interpretation (Stenning and van Lambalgen [2008]). It may be argued that higher results of group A than group B in ER stem just from a better grasp of meaning of logical constants corresponding to the natural language connectives employed in this test, in particular the conditionals. This also may be viewed as learning an important problem-solving ability. Nevertheless, we tried to minimize impact of this factor by formulating crucial premises as equivalent to biconditionals (see Appendix 2), by wording the instruction as evoking analytical rather than probabilistic interpretation of conditionals (cf. Singmann and Klauer [2011], Evans et al. [2010]) and by employing third-person perspective in solving all the tasks (cf. Beatty and Thompson [2012]). Nevertheless, possible influence of this factor on ER results cannot be dismissed.

## 5 Summary and conclusion

In this research we examined correlations between level of fluid intelligence (measured by Raven's Advanced Progressive Matrices test) and fluencies in two kinds of deductions: simple (operationalized by means of syllogistic reasoning tasks and measured by Deductive Reasoning test) and difficult ones (operationalized by means of erotetic reasoning and measured by Erotetic Reasoning test). Our subjects formed groups of lower (group C) and higher (groups A and B) abilities with respect to fluid intelligence and of extensive (group A) and very limited (groups B and C) training in formal logic. We observed that, although groups A and B performed better than group C in all three tests, group A obtained significantly higher results than group B in tasks involving difficult deductions, while their performance in Raven's and Deductive Reasoning tests were comparable. Group C obtained lower results in all three tests. We conclude on this basis, somewhat at odds with CHC model of intelligence, that fluency in difficult deductions, while related to fluid intelligence, depends also on subjects' experience and that this does not hold in case of simple deductions.

**Acknowledgments** Research for this paper was supported by the National Science Centre, Poland (DEC-2012/04/A/HS1/00715). We would like to thank Justyna Lembicz for her substantial assistance in carrying out this research and Monika Zawadzka for her assistance in translation of the tests materials (Appendices 1 and 2).

## References

Ajdukiewicz, K.: *Pragmatic Logic*. Dordrecht, Reidel (1974).

Baltes, P. B.: Theoretical Propositions of Life-Span Developmental Psychology: On the Dynamics between Growth and Decline. *Developmental Psychology*, 23:611–626 (1987).

Beatty, E. L., Thompson, V. A.: Effects of perspective and belief on analytic reasoning in a scientific reasoning task. *Thinking & Reasoning*, 18(4):441–460 (2012).

Benson, N., Hulac D. M. and Kranzler, J. H.: Independent Examination of the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV): What Does the WAIS-IV Measure? *Psychological Assessment* 22(1), 121–130 (2010).

Bergman Nutley, S., Söderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K. and Klingberg, T.: Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: a controlled, randomized study. *Developmental Science* 14(3), 591–601 (2011).

Blair, C.: Fluid Cognitive Abilities and General Intelligence: A Life-Span Neuroscience Perspective. In: Lerner and Overton [2010].

Cattell, R. B.: *Abilities: Their structure, growth, and action*. New York: Houghton Mifflin (1971).

Carroll, J. B: *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, MA: Cambridge University Press (1993).

Chooi, W.-T., Thompson, L. A.: Working memory training does not improve intelligence in healthy young adults. *Intelligence*, 40(6):531–542 (2012).

Dunn, O. J.: Multiple comparisons using rank sums. *Technometrics* 6:241–252 (1964).

Evans, J. St. B. T. In two minds: dual-process accounts of reasoning. *Trends in Cognitive Science*, 7(10):454–459 (2003)

Evans, J. St. B. T. Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning*, 18(1):5–31 (2012).

Evans, J. St. B. T., Handley, S., Neilens, H., Bacon, A. M., Over, D. E. The influence of cognitive ability and instructional set on causal conditional inference. *Quarterly Journal of Experimental Psychology*, 63, 892–909, (2010).

Gabbay, D. M., Woods, J.: *The Reach of Abduction. Insight and Trial.* Elsevier (2005).

Horn, J. L., and Noll, J.: Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Gensaft, & P. L. Harrison (eds.). *Contemporary intellectual assessment: Theories, tests, and issues*, NY: Guilford, 53–91 (1997).

Jaeggi, S. M., Buschkuehl, M., Jonides, J., Perrig, W. J.: Improving fluid intelligence with training on working memory. PNAS 105(19):6829–6833 (2008).

Karelitz, T. M., Jarvin, L., and Sternberg, R. J.: The Meaning of Wisdom and Its Development Throughout Life. In: Lerner and Overton [2010].

Klauer, K. J. and Willmes, K.: Inducing Inductive Reasoning: Does It Transfer to Fluid Intelligence? *Contemporary Educational Psychology* 27, 1–25 (2002).

Lerner, R. M., Overton, W. F. (eds.) *The Handbook of Life-Span Development. Cognition, Biology, and Methods*, John Wiley & Sons (2010).

Magnani, L. *Abductive Cognition. The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning.* Springer-Verlag, Berlin – Heidelberg (2009).

McArdle, J. J., Ferrer-Caja, E., Hamagami, F. and Woodcock, R. W.: Comparative Longitudinal Structural Analyses of the Growth and Decline of Multiple Intellectual Abilities Over the Life Span. *Developmental Psychology* 38(1), 115–142 (2002).

McGrew, K. S.: Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (eds.), *Contemporary intellectual assessment: Theories, tests, and issues* New York, NY: Guilford Press, 151–179 (1997).

McGrew, K. S.: The Cattell–Horn–Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed.), New York, NY: Guilford Press 136–181 (2005).

Oaksford, M., Chater, N. *Bayesian rationality: The probabilistic approach to human reasoning.* Oxford, UK: Oxford University Press (2007).

Paulewicz, B., Chuderski, A., Nęcka, E.: Insight problem solving, fluid intelligence, and executive control: A structural equation modeling approach. In: S. Vosniadou, D. Kayser, A. Protopapas (eds.), *Proceedings of the 2nd European Cognitive Science Conference*, Laurence Erlbaum, Hove (2007) 586–591.

Raven J., Raven, J. C., Court, J. H.: *Manual for Raven's Progressive Matrices and Vocabulary Scales* (Section 4: Advanced Progressive Matrices). San Antonio, TX: Harcourt Assessment (2003).

Ricco, R. B.: Development of Deductive Reasoning Across the Life Span. In: Lerner and Overton [2010].

Ricco, R. B., Overton, W. F.: Dual systems Competence $\leftarrow - \rightarrow$ Procedural processing: A relational developmental systems approach to reasoning. *Developmental Review* 31, 119–150 (2011).

Schneider, W. J. and McGrew, K. S.: The Cattell-Horn-Carroll model of intelligence. In D. Flanagan & P. Harrison (eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed.). New York: Guilford 99–144 (2012).

Singmann, H., Klauer, K. Ch.: Deductive and inductive conditional inferences: Two modes of reasoning. *Thinking & Reasoning*, 17(3):247–281 (2011).

Sloman, S. A. The empirical case for two systems of reasoning. *Psychological Bulletin*, 119:322 (1996).

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

Stanovich, K. E. Individual differences in reasoning and the algorithmic/intentional level distinction in cognitive science. In J. E. Adler & L. J. Rips (Eds.), *Reasoning: Studies of human inference and its foundations* (pp. 414436). Cambridge: Cambridge University Press (2008).

Stanovich, K. E., West R. F.: Individual differences in reasoning: Implications for the rationality debate. *Behav. Brain Sci.* 23, 645–726.

Stenning, K., van Lambalgen, M.: *Human Reasoning.* The MIT Press (2008).

Sternberg, R. J.: Increasing fluid intelligence is possible after all. PNAS 105(19):6791–6792, (2008).

Sternberg, R. J. Intelligence. In K. J. Holyoak & R. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 751–773) New York: Cambridge University Press (2005).

Tabachnick, B.G., Fidell, L.S.: *Using Multivariate Statistics* (5th edition). Pearson/Allyn & Bacon (2007).

Urbański, M., Łupkowski, P.: Erotetic Search Scenarios: Revealing Interrogator's Hidden Agenda. In: P. Łupkowski, M. Purver (eds.), Semantics and Pragmatics of Dialogue, Polskie Towarzystwo Kognitywistyczne, Poznań (2010) 67–74.

Wagner-Egger, P.: Conditional reasoning and the Wason selection task: Biconditional interpretation instead of reasoning bias. *Thinking & Reasoning*, 13(4):484–505 (2007).

Ward, S. L., Overton, W. F.: Semantic familiarity, relevance, and the development of deductive reasoning. *Developmental Psychology*, 26, 488–493 (1990).

Wechsler, D.: *Wechsler Adult Intelligence Scale – Fourth Edition.* San Antonio, TX: Pearson Assessment, (2008).

Wilhelm, O.: Measuring Reasoning Ability. In: Wilhelm and Engle [2005], 373–392.

Wilhelm, O., Engle, R. W. (eds.): *Handbook of Understanding and Measuring Intelligence.* Sage Publications, Thousand Oaks, CA (2005).

Wiśniewski, A.: *The Posing of Questions. Logical Foundations of Erotetic Inferences.* Dordrecht: Kluwer AP (1995).

Wiśniewski, A.: *Questions, Inferences, and Scenarios.* London: College Publications (2013).

## Appendix 1: English translation of tasks from Deductive Reasoning Test

Imagine playing the game in whose realm there exist various peculiar objects: CHIGS, GIDGITS, OZACKS and many others. Though you do not know what these objects actually are, you realize that the realm of the game is controlled by certain rules and conditions, just like in the real world. Answer the questions in each task below, using the information given in the two statements above the line. Mark the right answer in the right box.

1. Every TWAIL is an ANKH.
   Every ANKH is a GIMBEAT.
   Is every TWAIL a GIMBEAT ?

   ☐ Definitely yes ☐ Definitely no ☐ It cannot be stated

2. No CHIG is a DAZZLA.
   No DAZZLA is a GOURD.
   Is any GOURD a CHIG ?

   ☐ Definitely yes ☐ Definitely no ☐ It cannot be stated

3. Every SULERL is a GIDGIT
   No GIDGIT is a LURKIE.
   Is every LURKIE a SULERL ?

   ☐ Definitely yes ☐ Definitely no ☐ It cannot be stated

4. MUZACK is shorter than CHIMAR.
   CHIMAR is shorter than OZACK.
   Is MUZACK shorter than OZACK ?

   ☐ Definitely yes ☐ Definitely no ☐ It cannot be stated

5. FUNKER is taller than CHULO.
   CHULO is taller than PIGGA.
   s PIGGA shorter than FUNKER ?

   ☐ Definitely yes ☐ Definitely no ☐ It cannot be stated

6. BRINJAL is smaller than THEREO.
   THEREO is bigger than GHURK.
   Is BRINJAL bigger than GHURK ?

   ☐ Definitely yes ☐ Definitely no ☐ It cannot be stated

## Appendix 2: English translation of an exemplary task from Erotetic Reasoning Test

In the capital of a certain country someone planted a bomb in the palace of the king. The best royal engineer, who arrived immediately, established the following evidence:

1. There are three wires in the bomb: green, red and orange.
2. To disarm the bomb either the green or the red wire must be cut. Cutting the wrong wire will cause an explosion.
3. If the bomb has been planted by Steve, cutting the green wire will disarm it.
4. If the bomb has been planted by John, cutting the red wire will disarm it. Moreover, no one but John would have used the red wire.
5. If the bomb has not been planted on an even day of the month, the culprit is Steve.
6. The bomb has been planted by Steve, or by John, or by someone else.

Each of the following questions below can be answered either ‘yes’ or ‘no’. Mark the question to which the answer (regardless of it being ‘yes’ or ‘no’) will allow you to establish in the shortest possible time which wire should be cut in order to disarm the bomb:

[ ] Was the bomb planted on an even day of the month?
[ ] Was the bomb planted by Steve?
[ ] Was the bomb planted by John?
[ ] Was the bomb planted by someone else than Steve or John?

Justify your choice.