Peirce: an Algorithm for Abductive Reasoning Operating with a Quaternary Reasoning Framework

Felipe Rodrigues, Carlos Eduardo A. Oliveira, and Osvaldo Luiz de Oliveira

FACCAMP, Campo Limpo Paulista (SP),
Brazil

rodrigues_felipe7@hotmail.com, carlos.br@gmail.com, osvaldo@faccamp.br

Abstract. Abductive reasoning algorithms formulate possible hypotheses to explain observed facts using a theory as the basis. These algorithms have been applied to various domains such as diagnosis, planning and interpretation. In general, algorithms for abductive reasoning based on logic present the following disadvantages: (1) they do not allow the explicit declaration of conditions that may affect the reasoning, such as intention, context and belief; (2) they allow little or no consideration for criteria required to select good hypotheses. Using Propositional Logic as its foundation, this study proposes the algorithm Peirce, which operates with a framework that allows one to explicitly include conditions to conduct abductive reasoning and uses a criterion to select good hypotheses that employs metrics to define the explanatory power and complexity of the hypotheses. Experimental results suggest that abductive reasoning performed by humans has the tendency to coincide with the solutions computed by the algorithm Peirce.

Keywords: Abductive reasoning, automated Reasoning, logic, human factors.

1 Introduction

Abductive reasoning formulates hypotheses to explain observed facts using a theory as the basis. Numerous intellectual tasks make use of abductive reasoning, including medical diagnostics, fault diagnostics, scientific discovery, legal argumentation and interpretation.

Abductive reasoning algorithms based on logic frequently operate with the Theory, Hypotheses and Facts (THF) ternary reasoning framework (as shown in [2], [4], [5] and [11]). When these algorithms are formalized using Propositional Logic (PL) [9], the THF framework is frequently instantiated according to Definition 1.

Definition 1 (THF reasoning framework). The THF reasoning framework for abductive reasoning is a system ⟨T, H, F⟩ that consists of the following:

- A finite and non-empty theory set, T = { t_1, t_2, t_3, ..., t_m }, of PL sentences denoting $t_1 \wedge t_2 \wedge t_3 \wedge ... \wedge t_m$. This set represents the hypotheses that must be assumed as truth during the reasoning process.

- A finite hypotheses set, H = { h_1, h_2, h_3, ..., h_n }, of PL sentences denoting $h_1 \vee h_2 \vee h_3 \vee \ldots \vee h_n$. This set represents the hypotheses that along with the set T explain the facts represented by the set F.
- A set with a single fact, F = { f }, where f is a PL literal (atom or negated atom). This set represents an occurrence of an evidence, a manifestation, a symptom, an observation, a mark or a sign to be explained through abductive reasoning.

Having the T and F sets as input, an abductive reasoning algorithm should find a set of hypotheses H that satisfies the following conditions:

$$T \nvDash F, \tag{1}$$

$$T \cup \{ h \} \vDash F, \forall h \in H, \tag{2}$$

$$T \cup \{ h \} \nvDash \bot, \forall h \in H, \tag{3}$$

$$\{ h \} \nvDash F, \forall h \in H. \tag{4}$$

The statements above refer to the concept of logical consequence described in Definition 2.

Definition 2 (A ⊨ B, i.e., B is a logical consequence of A). Let A = { a_1, a_2, a_3, ..., a_n } and B = { b_1, b_2, b_3, ..., b_m } be two finite and non-empty sets of PL sentences. Then, A ⊨ B if and only if the interpretations in which $a_1 \wedge a_2 \wedge a_3 \wedge \ldots \wedge a_n$ is true, $b_1 \wedge b_2 \wedge b_3 \wedge \ldots \wedge b_m$ is also true.

The condition (1) prevents that the theory set T alone has as logical consequence the facts set F. Hypotheses satisfying condition (2) are called **candidate hypotheses**, and they can explain the single fact denoted by F. Candidate hypotheses satisfying condition (3) are called **consistent hypotheses**. Conversely, candidate hypotheses that do not satisfy condition (3) are called inconsistent hypotheses and should be discarded. Candidate hypotheses that satisfy condition (4) are called **explanatory hypotheses**. Conversely, candidate hypotheses that do not satisfy condition (4) are called non-explanatory hypotheses and should be discarded.

Example 1. Joseph has a large lawn in front of his house. One day, Joseph arrives at home and observes that the lawn is wet. Considering only that (1) rain can make the lawn wet and that (2) sprinklers installed across the lawn can make it wet, which hypotheses can explain the fact that the lawn is wet?

One possible formalization using the THF framework consists in defining:

- Propositions 'r: **Rain** occurred', 's: **Sprinklers** were activated' and 'w: Lawn is **wet**'.
- A theory set T = { $r \rightarrow w$, $s \rightarrow w$ }.
- A fact set F = { w }.

The theory set T and the fact set F satisfy condition (1), whereas a theory T_1 = { $r \rightarrow w$, $s \rightarrow w$, w } has F as its logical consequence; therefore T_1 and F do not satisfy condition (1). Let H = { r, s, $r \wedge s$, $r \wedge \neg w$, w } be a set of candidate hypotheses. Each hypothesis $h \in H$ satisfies condition (2), and each hypothesis $h \in$ { r, s, $r \wedge s$, w } satisfies condition (3), i.e., they are consistent. However, the hypothesis $r \wedge \neg w$ is inconsistent because T ∪ { $r \wedge \neg w$ } ⊨ ⊥; therefore, it must be discarded. Each

hypothesis $h \in \{ r, s, r \wedge s \}$ satisfies condition (4); however, the hypothesis w is not explanatory because $\{ w \} \models$ F; therefore, it must also be discarded. Thus, removing the inconsistent and non-explanatory hypotheses from H, we obtain H = $\{ r, s, r \wedge s \}$.

In general, abductive algorithms work as follows: having the theory set T and the fact set F as the input, the algorithm verifies whether or not the condition (1) has been satisfied; if the condition (1) is not satisfied, then there are no hypotheses to be formulated because F is a logical consequence of T; however, if the condition (1) is satisfied, then the algorithm formulates a finite set of possible hypotheses H that satisfies the condition (2). Next, the algorithm removes from H the hypotheses that do not satisfy the conditions (3) and (4), and thus returning the resulting set H as an answer by the algorithm.

Some algorithms, however, include an additional step with the goal of letting in H only hypotheses considered good, according to extra-logical criteria. A criterion commonly used is "simplicity", which considers, for example, an atomic hypothesis better than a composite hypothesis, e.g., r is better than $r \wedge s$.

The need to represent conditions such as context, circumstance and intention is common and important when conducting abductive reasoning. For example, reasoning to make a medical diagnosis considering the context of diseases of a region. Operating with a THF reasoning framework, the existing algorithms to perform abductive reasoning have the disadvantage of forcing the representation of these conditions in the theory set T. This solution is not appropriate because representing conditions in the theory set T mischaracterizes the theory, making it less general and more *ad hoc* (specific to explain what one wants to explain).

Abductive reasoning formulates hypotheses, and some of these hypotheses may be better at explaining the facts than others. Today we do not know, exactly, which criteria determine what makes a hypothesis better than another, authors from several fields [3] [8] [10] [13] [16] [17] have suggested that abductive reasoning involves the selection of good hypotheses. However, the existing abductive reasoning algorithms have the disadvantage of dedicating little or no consideration for criteria required to select good hypotheses.

Many practical applications of reasoning require the definition of a set of $n \geq 2$ facts. However, the many existing algorithms have the disadvantage of operating with only a single fact.

This work proposes an algorithm, called Peirce, that performs abductive reasoning, and this algorithm differs from the existing solutions mainly because (1) it works with a reasoning framework called TCHF (Theory, accepted Conditions, Hypotheses and Facts), thus allowing conditions to be explicitly represented; (2) it allows $n \geq 2$ facts to be represented; and (3) it introduces a criterion to select good hypotheses that employ metrics to define the explanatory power and the complexity of the hypotheses.

Section 2 describes the algorithm Peirce, dedicating particular attention to the design and operation of the TCHF reasoning framework (Subsection 2.1) and the definition of a criterion to select good abductive hypotheses (Subsection 2.2). The pseudocode for the algorithm Peirce is presented and discussed in Subsection 2.3. Section 3 details an experimental study conducted to verify whether the solutions computed by the algorithm Peirce tend to coincide with the abductive reasoning

performed by humans. Section 4 describes related works, highlighting the differences with this work. Section 5 presents the conclusions.

2 The Algorithm Peirce

The abductive reasoning algorithm proposed in this study has been named Peirce in honor of the American philosopher Charles Sanders Peirce, who created the concept of abductive reasoning [14]. The following subsections detail the reasoning framework used by the algorithm Peirce, a criteria to select good hypotheses and the pseudocode of the algorithm.

2.1 TCHF Reasoning Framework

The TCHF reasoning framework proposed in this study differs from the classic THF reasoning framework (Definition 1) by including the accepted conditions set C and by redefining the facts set F to allow the declaration of not just one single fact, but rather a finite number of one or more facts. The TCHF framework is formalized in Definition 3 and uses the PL sentences in HF form as specified in Definition 4.

Definition 3 (TCHF reasoning framework). The TCHF framework for abductive reasoning is a system ⟨T, C, H, F⟩ consisting of the following:

- A finite and non-empty theory set, T = { t_1, t_2, t_3, ..., t_m }, of PL sentences in HF form denoting $t_1 \wedge t_2 \wedge t_3 \wedge ... \wedge t_m$. This set represents the hypotheses that must be assumed as truth during the reasoning process.
- A finite hypotheses set, H = { h_1, h_2, h_3, ..., h_n }, of PL sentences in HF form denoting $h_1 \vee h_2 \vee h_3 \vee ... \vee h_n$. This set represents the hypotheses that along with the sets T and C explain the facts represented by the set F.
- A finite accepted conditions set, C = { c_1, c_2, c_3, ..., c_p }, of PL sentences in HF form denoting $c_1 \wedge c_2 \wedge c_3 \wedge ... \wedge c_p$. This set represents the conditions that must be assumed as truth during the reasoning process.
- A finite and non-empty facts set F = { f_1, f_2, f_3, ..., f_q } of PL positive literals, denoting $f_1 \wedge f_2 \wedge f_3 \wedge ... \wedge f_q$. The role of this set is to represent evidences, manifestations, symptoms, observations, marks or signs to be explained by the abductive reasoning.

Definition 4 (HF form). A sentence of PL in the HF form is an acyclic sentence written in one of the following formats:

- $a_1 \wedge a_2 \wedge a_3 \wedge ... \wedge a_n$, where a_i $(1 \leq i \leq n)$ are literals.
- $a_1 \vee a_2 \vee a_3 \vee ... \vee a_n$, where a_i $(1 \leq i \leq n)$ are negative literals.
- $a_1 \wedge a_2 \wedge a_3 \wedge ... \wedge a_n \rightarrow b_1 \wedge b_2 \wedge b_3 \wedge ... \wedge b_m$, where a_i $(1 \leq i \leq n)$ are positive literals and b_j $(1 \leq j \leq m)$ are literals.
- $a_1 \vee a_2 \vee a_3 \vee ... \vee a_n \rightarrow b_1 \vee b_2 \vee b_3 \vee ... \vee b_m$, where a_i $(1 \leq i \leq n)$ are literals and b_j $(1 \leq j \leq m)$ are negative literals.

The restriction of TCHF framework to sentences in the HF form aims to make the algorithm Peirce run the conversion of sentences in polynomial time, because sentences in HF form can be easily converted into Horn Clauses [9] (disjunction of literals with at most one positive literal). As will be described in Subsection 2.3, algorithm Peirce uses Resolution as the inference mechanism and this mechanism can be efficiently implemented on Conjunctive Normal Form (CNF) sentences with Horn Clauses.

The set C gives the TCHF reasoning framework the advantage of allowing the explicit definition of conditions that in the classical THF framework, would normally be declared within the theory set T. Thus, the set C avoids "contaminating" the set T with sentences that fundamentally do not belong to the theory. Moreover, this makes it easier to represent two or more instances of abductive reasoning that share the declarations of T and F but differ in the set of accepted conditions. Example 2, which is described next, illustrates the use of the TCHF reasoning framework.

Example 2. Consider once more the scenario described in Example 1 in which Joseph arrives at home and observes his lawn wet. However, let us say that Joseph knows that the water tank supplying the sprinklers has been empty for a month; therefore, under this condition, the sprinklers could not have been activated.

One possible formalization using the TCHF framework is to define the following:

- Propositions '*r*: **Rain** occurred', '*s*: **Sprinklers** were activated', '*w*: Lawn is **wet**' and '*t*: Water **tank** that supplies sprinklers is empty'.
- A theory T = { $r \rightarrow w$, $s \rightarrow w$ }, which is the same as in Example 1.
- A set of accepted conditions C = { t, $t \rightarrow \neg s$ }.
- A set of facts F = { w }, which is the same as in Example 1.

The conditions in abductive reasoning are motivated by several factors which are linked to context (information associated to space), circumstances (information associated with time), intention (manifestation of the will to reach some wanted conclusions), belief or faith (information that is accepted on principle) etc. Examples of specific conditions used in abductive reasoning are as follows: (1) In abductive reasoning used for medical diagnoses, regional context may allow one to specify a set of diseases that are common or uncommon for a given region; (2) In abductive reasoning used for anthropological studies, the specification of possible agents that might have been responsible for the death of a hominid based on the knowledge that the hominid lived 4 million years ago (circumstance); (3) In abductive reasoning for judicial decisions, possible conditions may be specified with the intent of acquitting (or condemning) a defendant; and (4) In abductive reasoning for religious or metaphysical beliefs, the faith or belief that there is life after death can be declared as a condition upon which reasoning are made.

Taking the sets T and F as inputs, an abductive reasoning algorithm operating with the TCHF framework should find a set of hypotheses H that satisfies the following conditions:

$$T \cup C \not\models F, \tag{5}$$

$$T \cup C \cup \{ h \} \models_p F, \forall h \in H, \tag{6}$$

$$T \cup C \cup \{ h \} \nvDash \bot, \forall h \in H, \quad (7)$$

The condition (6) uses the partial logical consequence as defined in Definition 5.

Definition 5 ($A \vDash_p B$, i.e., B is a partial logical consequence of A). Let $A = \{ a_1, a_2, a_3, ..., a_n \}$, $B = \{ b_1, b_2, b_3, ..., b_m \}$ and $C = \{ c_1, c_2, c_3, ..., c_q \}$, $C \subseteq B$, be three finite and non-empty sets of PL sentences. Then, $A \vDash_p B$ if only if the interpretations in which $a_1 \wedge a_2 \wedge a_3 \wedge ... \wedge a_n$ is true, $c_1 \wedge c_2 \wedge c_3 \wedge ... \wedge c_q$ is also true.

2.2 Selection of Good Abductive Hypotheses

In general, several hypotheses may be able to explain observed facts. However, certain hypotheses may explain facts better than others. Therefore, abductive reasoning can be observed as a process that formulates $m \geq 1$ general hypotheses followed by the selection of $n \leq m$ good hypotheses. Naturally, selection criteria must be established, but it is still difficult to define the conditions that make a hypothesis good.

Contemporary philosophers have analyzed the issue of selecting good hypotheses. Harman [8] considers abduction to be an inference of the best explanation and argues that the best hypothesis is the simplest, most plausible and is the least *ad hoc*. By comparing theories (e.g., Darwin's Theory of Evolution vs. Creationist Theory or Lavoisier's Theory of Combustion vs. Phlogiston Theory), Thagard [16, 17] establish criteria that explain the preference for one hypothesis over another and considers the best hypothesis to be the most consilient (explains more facts), the most simple, and it would provide the best analogy with hypotheses that explain facts in other domains.

Criteria to select good hypotheses have been extensively studied in the fields of philosophy (e.g., [2] [8] [16]), psychology (e.g., [13]) and artificial intelligence (e.g., [3] [10] [15]). However, the precise formulation of these criteria remains controversial. In general, factors such as the "explanatory power" and the "complexity" of a hypothesis are recurrent and have similar connotations across several studies. Therefore, this study has proposed using these two factors to develop a selection criterion. Aiming at the development of algorithms to perform abduction that need dealing with quantitative measures for the explanatory power and the complexity of a hypothesis, this study proposes an understanding of these factors as follows:

- Explanatory power (or comprehensiveness): the explanatory power of a hypothesis quantifies the degree to which it is capable of explaining the facts involved in the reasoning. A metric for a hypothesis' explanatory power is given by the ratio between the number of facts it can explain and the total number of facts to be explained by the abductive reasoning process. For example, a hypothesis that explains 4 out of 5 facts has an explanatory power of 4/5, and a hypothesis that explains all of the facts has an explanatory power of 1.
- Complexity: the complexity factor refers to how many different elements and relationships are present in a hypothesis. A metric for hypothesis complexity is the number of atomic propositions that it contains. For example, hypothesis r has a complexity of 1, and hypothesis $r \wedge s \wedge w$ has a complexity of 3.

Based on the metrics for explanatory power and complexity, this study proposes a criterion to select good hypotheses, which is declared in Definition 6.

Definition 6 (A criterion to select good hypotheses). Given a set H of candidate hypotheses to explain a set F of facts, $h \in H$ is considered a good hypothesis if it satisfies all the following conditions:

- The explanatory power of h is equal to or greater than a constant λ_1. The constant $\lambda_1 = 0.5$ has been used in the experiments described in this article.
- The complexity of h is equal to or less than a constant λ_2. The constant[1] $\lambda_2 = 5$ has been used in the experiments described in this article.
- The hypothesis h has the minimum complexity among all of the hypotheses that have the maximum explanatory power in H.

Examples 3 and 4 illustrate the application of Definition 6.

Example 3. Diseases manifest themselves through symptoms. Consider the following:

- Propositions '*c*: Disease is **cold'**, '*p*: Disease is **pneumonia'**, '*r*: Disease is **rhinitis'**, '*f*: Symptom is **fever'**, '*h*: Symptom is **headache'**, and '*z*: Symptom is **coryza'**;
- Theory T = { $p \rightarrow f \wedge z \wedge h$, $c \rightarrow f \wedge z$, $r \rightarrow h \wedge z$ }, the empty set C = { } of accepted conditions and observed facts set F = { *f*, *z*, *h* } (symptoms);
- A set of candidate hypotheses H = { *p*, *c*, *r*, $p \wedge c$, $p \wedge r$, $c \wedge r$, $p \wedge c \wedge r$ }.

Table 1 describes the explained facts, explanatory power and complexity of each candidate hypothesis $h \in H$.

Table 1. Explained facts, explanatory power and complexity of candidate hypothesis of the Example 3. The '√' signals an explained fact.

Hypothesis	Explained facts			Explanatory power	Complexity
	f	***z***	***h***		
p	√	√	√	1	1
c	√	√		0.66	1
r		√	√	0.66	1
$p \wedge c$	√	√	√	1	2
$p \wedge r$	√	√	√	1	2
$c \wedge r$	√	√	√	1	2
$p \wedge c \wedge r$	√	√	√	1	3

All of the hypotheses have an explanatory power equal to or greater than $\lambda_1 = 0.5$ and complexity equal to or less than $\lambda_2 = 5$. The hypotheses p, $p \wedge c$, $p \wedge r$, $c \wedge r$, $p \wedge c \wedge r$ have explanatory power equal to 1, which is the maximum among all candidate

[1] The λ_1 and λ_2 values were chosen to coincide with human factors. Considering Miller experiment [12], human memory and human processing capacity is limited to 7 ± 2 simultaneous elements, hence $\lambda_2 = 5$. Good hypotheses explain at least 50% of the facts, hence $\lambda_1 = 0.5$.

hypotheses. Among these hypotheses with maximum explanatory power, hypothesis *p* has the complexity equals to 1, which is the minimum among the hypotheses. Therefore, *p* is a good hypothesis according to Definition 6.

Example 4. Another example involving diseases and symptoms. Consider the following:

- Propositions '*d*: Disease is **dengue**', '*u*: Disease is **flu**', '*b*: Symptom is **breathlessness**', '*f:* Symptom is **fever**', '*h*: Symptom is **headache**', '*m*: Symptom is **muscle pain**', '*r*: Symptom is **red spots**' and '*s*: Symptom is **sneezing**';
- Theory T = { $u \rightarrow f \wedge h \wedge m \wedge$ s, $d \rightarrow f \wedge h \wedge m \wedge$ r }, the empty set C = { } of accepted conditions and observed facts set F = { *f, h, m, b* } (symptoms);
- A set H = { *u, d,* $u \wedge d$ } of candidate hypotheses.

Table 2 describes the explained facts, explanatory power and complexity of each candidate hypothesis $h \in$ H.

Table 2. Explained facts, explanatory power and complexity of candidate hypothesis of the Example 4. The '√' signals an explained fact.

Hypothesis	Explained facts				Explanatory power	Complexity
	f	***h***	***m***	***b***		
U	√	√	√		0.75	1
D	√	√	√		0.75	1
$u \wedge d$	√	√	√		0.75	2

All of the hypotheses have an explanatory power equals to 0.75 (i.e., explanatory power equal to or greater than $\lambda_1 = 0.5$) and complexity equal to or less than $\lambda_2 = 5$. Hypotheses *u* and *d* have a complexity of 1, which is the minimum across all of the candidates. Therefore, *u* and *d* are good hypotheses according to Definition 6.

2.3 Pseudocode for the Algorithm Peirce

Figure 1 presents the pseudocode for the algorithm Peirce. The algorithm Peirce formulates hypotheses that comply with equations (5), (6), (7) and the criterion to select good hypotheses of the Definition 6. Synthetically, the algorithm Peirce formulates candidate hypotheses and stores them in set H (line 11). Next, the algorithm removes inconsistent hypotheses from H (line 12) and then selects and leaves only the good hypotheses in H (line 13). The details of the algorithm are described below.

The algorithm uses the Resolution rule of inference for PL sentences in CNF expressed with Horn Clauses [9]. Candidate hypotheses are hypotheses *h* that satisfy equation (6). To compute these hypotheses, the algorithm translates the set of sentences T ∪ C ∪ ¬F to CNF expressed with Horn Clauses (line 4) and applies the mechanism of resolution (line 5). The result of the resolution is stored in the data structure R (set of clauses). If R contains at least one empty clause, then T ∪ C ⊨ F and no hypotheses are formulated (lines 6 and 7). If R does not contain an empty

clause, then T ∪ C ⊭ F, equation (5) is met and candidate hypotheses can be formulated.

At line 11 each clause in R presents the possibility of formulating a hypothesis. Because R is in CNF, negating each clause results in a candidate hypothesis. The algorithm `Formulate_Candidate_Hypotheses` (line 11), operates as follows: (1) the algorithm negates each of the $m \geq 1$ clauses in R to obtain m first-candidates hypotheses h_1, h_2, h_3, ..., h_m, (2) combines these m first-candidates hypotheses in pairs to obtain conjunctive hypotheses of the type $h_i \wedge h_j$ ($i \neq j$); and (3) combines the m first-candidates hypotheses three by three to obtain conjunctive hypotheses of the type $h_i \wedge h_j \wedge h_k$ ($i \neq j \neq k$); … and then combines the m first-candidates hypotheses q by q to obtain conjunctive hypotheses of the type $h_i \wedge h_j \wedge ... \wedge h_q$ ($i \neq j \neq ... \neq q$), where q = min (m, λ_2) and is λ_2 the constant that defines the maximum complexity of the hypotheses (λ_2 is defined in Definition 6).

```
    Algorithm Peirce(T, C, F)
    Input
    Theory set T, accepted condition set C and facts set F
    (specification is given in Definition 3).
    Output
    Hypotheses set H (specification is given in Definition 3).
1   {
2     if Consistent(T, C) then
3     {
4       R := Conjunctive_Normal_Form_Horn_Clauses(T, C, ¬F);
5       R := Resolution(R);
6       if R contains an empty clause then
7         write ("No hypotheses to formulate: T ∪ C ⊨ F");
9       else
10      {
11        H := Formulate_Candidate_Hypotheses(R);
12        H := Remove_Inconsistent_Hypotheses(T, C, H);
13        H := Select_Good_Hypotheses(T, C, H, F);
14      }
15    }
16    else
17      write("Unable to formulate hypotheses: T ∪ C ⊨ ⊥.");
18  }
```

Fig. 1. Algorithm Peirce.

At line 12, the algorithm `Remove_Inconsistent_Hypotheses` receives a set H of candidate hypotheses and removes from H hypotheses that do not satisfy T ∪ C ∪ { h } ⊭ ⊥ (conformity to equation (7)). The algorithm works as follows: For each $h \in$ H: (1) the algorithm translates the sentences in the set T ∪ C ∪ { h } to CNF expressed with Horn Clauses, (2) applies the Resolution mechanism to this system of

sentences and (3) removes hypothesis *h* from H if the Resolution mechanism derives an empty clause.

The algorithm `Select_Good_Hypotheses` receives a set H of candidate hypotheses that are all consistent and then operates as follows: (1) it computes the explanatory power and complexity of each hypothesis $h \in$ H, (2) removes all of the hypotheses *h* with explanatory power below some constant λ_1 (0.5 in our experiments) or complexity above some constant λ_2 (5 in our experiments), (3) computes set E with the hypotheses that have the maximum explanatory power in H, (4) computes set X with the hypotheses that have the minimum complexity in E and (5) returns set X as answer.

It can be proved that the Peirce algorithm computes three different types of solutions: (1) H = F when the theory and the accepted conditions does not allow Peirce algorithm to formulate explanatory hypotheses; (2) H = { } when Peirce algorithm does not consider any hypotheses to be good, among the candidate hypotheses; (3) H contains at least one explanatory hypothesis; In this last type, H does not contain non-explanatory hypothesis.

Example 5 illustrates a run of the algorithm Peirce.

Example 5. This example illustrates the execution of the algorithm Peirce using the scenario and formalization from Example 2. Therefore, the algorithm Peirce receives as input the theory T = { $r \rightarrow w$, $s \rightarrow w$ }, the set of accepted conditions C = { t, $t \rightarrow \neg s$ } and the set of facts F = { w }. Because T ∪ C ⊭ ⊥, algorithm `Consistent(T, C)` returns the value true (line 2), and the data structure R is filled in with T ∪ C ∪ ¬F in CNF expressed with Horn Clauses. The following is then established:

- At line 4: R = { { $\neg r$, w }, { $\neg s$, w }, { t }, { $\neg t$, $\neg s$ }, { $\neg w$ } };
- At line 5 after Resolution: R = { { $\neg r$ }, { $\neg s$ } }.

Because there are no empty clauses in R (test at line 6), the candidate hypotheses are formulated at line 11. Thus, H = { r, s, $r \wedge s$ } at line 11 after executing `Formulate_Candidate_Hypotheses`. Because T ∪ C ∪ { s } ⊨ ⊥ and T ∪ C ∪ { $r \wedge s$ } ⊨ ⊥, hypotheses s and $r \wedge s$ are removed from H by the algorithm `Remove_Inconsistent_Hypotheses` (line 12), leaving H = { r }. The hypothesis r has an explanatory power of 1, a complexity of 1 and the minimum complexity of all hypotheses with maximum explanatory power in H (r is the only hypothesis in H), therefore the algorithm `Select_Good_Hypotheses` (line 13) selects r as a good hypothesis. The algorithm Peirce thus returns as answer H = { r }.

In general, the complexity of logic-based abduction is NP-complete [6]. However, the algorithm Peirce has a running time $O(n^{2+\lambda_2})$. As λ_2 is a constant, typically equals to 5, Pierce algorithm runs in polynomial time. This occurs by the following facts. The algorithm `Conjunctive_Normal_Form_Horn_Clauses` has running time $O(n)$ because since every sentence of T, C and F is restricted to HF form (Definition 4) they can be transformed directly into Horn Clauses in $O(1)$. The execution of the Resolution mechanism of the PL sentences in CNF with Horn Clauses can be done in $O(n^2)$. Thus, `Consistent` and `Resolution` algorithms have running time $O(n^2)$. The algorithm `Formulate_Candidate_Hypotheses` has a running time of $O(n^{\lambda_2})$ because produces at most hypotheses combinations $O(n^2) + O(n^3) + \ldots + O(n^{\lambda_2})$. The

algorithm `Remove_Inconsistent_Hypotheses` has a running time of $O(n^{2+\lambda_2})$ because executes at most a constant amount of $O(n^{\lambda_2})$ resolutions each of them in $O(n^2)$. The algorithm `Select_Good_Hypotheses` has a running time of $O(n \log n)$, to sort and select the set of hypotheses with minimal complexity among the hypotheses with maximum explanatory power.

3 Tendency of Solutions Computed by the Algorithm Peirce to Coincide with Abductive Reasoning Done by Humans

A study was realized to verify whether the abductive reasoning performed by humans tends to coincide with the solutions computed by the algorithm Peirce. The study was conducted using a questionnaire containing ten questions, with each question presenting an implicit description of a theory, observed facts and accepted conditions. The alternatives for each question present possible abductive hypotheses. Table 3 illustrates in the left column one question in the questionnaire.

Table 3. Example of a question used in the questionnaire. The left column describes the question itself, and the right column presents the corresponding formalization to the question and solution as computed by the algorithm Peirce.

Question	**Formalization and solution computed by the algorithm Peirce**
Joshua is in the desert and sees something green in the distance. What would best explain what Joshua sees? a) I am convinced that it is a lawn. b) I am convinced that it is a cactus. c) I am convinced that it is a green flag. d) It could be either a cactus or a green flag.	**Propositions**: '*c*: It is a **cactus**', '***d***: It is a **desert**', '*f*: It is a green **flag**', '*l*: It is a **lawn**', '*s*: Joshua **sees** something green'. T = { $l \rightarrow s$, $c \rightarrow s$, $f \rightarrow s$ }, C = { d, $d \rightarrow \neg l$ }, F = { s }. **Solution** - After formulating candidate hypotheses (line 11): H = { l, c, f, $l \wedge c$, $l \wedge f$, $c \wedge f$, $l \wedge c \wedge f$ }. - After removing inconsistent hypotheses (line 12): H = { c, f, $c \wedge f$ }. - After selecting good hypotheses (line 13): H = { c, f }, i.e., the alternative 'd' coincides with the solution of algorithm Peirce.

The questionnaire, validated by a pilot-test with 25 individuals, was designed to be answered in 15 minutes. A total of 133 undergraduate and graduate students participated in the study. The profile of the participants showed a slight predominance of female individuals (53%) and ages ranging from 18 to 60 years, with an average and median close to 25 years.

Each participant's answers to the questionnaire were computed, and one point was attributed to each answer on the questionnaire that coincided with a solution produced by the algorithm Peirce. The results showed an average of 86 answers coinciding with the algorithm Peirce and 47 that did not coincide.

The Chi-square (χ^2) test at 1% significance was used as a statistical measure of the significance with which the participants' answers coincided with solutions produced by the algorithm Peirce. For the studied population, the χ^2 test suggested that the coincidence between the participants' answers and the solutions computed by the algorithm Peirce was significant: (χ^2 (1) = 11.44, *p*–value = 0.001 < 0.01).

4 Related Works

Different approaches have been used to develop algorithms for abductive reasoning. Among the many contributions, there are proposals that use search techniques [15] and probabilistic reasoning over Bayesian Networks [7]. Logic approaches are based on two types of contributions: (1) proposal of new algorithms and (2) extension of traditional logical programming to process abductive reasoning problems.

Examples of type 1 contributions include [2] and [5]. Both of the proposals refer to abductive reasoning algorithms that operate with a THF reasoning framework (Definition 1). The main differences between these proposals and those of the present study are as follows: (1) They allow only one fact to be declared; (2) They do not allow define explicitly a set of accepted conditions; and (3) Semantic Tableaux is used in the proposal described in [2] instead of Resolution as the mechanism of inference.

Contributions of type 2 include Abductive Logic Programming (ALP) [11] and use the languages Prolog with Constraint Handling Rules (CHR) [1] [4]. The main differences between these proposals and those of the present study are as follows: (1) They operate with Predicate Logic; (2) They require special "abducible" predicates (possible hypotheses) to be declared; and (3) They dedicate little attention to criteria to select good abductive hypotheses.

Studies related to the one presented here, that address the selection of good hypotheses, include [8], [16] and, recently, [3]. This work differs from proposals [8] and [16] mainly by the proposed metrics for complexity and explanatory power of hypotheses.

5 Conclusions

The abductive reasoning algorithm Peirce is distinct from other solutions mainly because it employs the TCHF reasoning framework and a simple criterion for selecting good hypotheses that consider quantitative metrics to define the explanatory power and complexity of the formulated hypotheses.

The TCHF reasoning framework has shown itself to be useful in organizing the elements that participate in abductive reasoning because it does not "contaminate" the theory with sentences that fundamentally do not belong to the theory. This framework provides an additional advantage because it explicitly exposes the conditions (contexts, circumstances, intentions etc.) under which the reasoning process is conducted, which is fundamental and frequent in the formulation of abductive reasoning.

The criteria for selecting good hypotheses are subjects of ongoing research. There is no consensus as to which criteria should be used and under which circumstances or in which domains they work. The criterion used by the algorithm Peirce, which is described in Definition 6, attempt to produce a simple algorithm that works in practice. Alternatives to Definition 6 exist and can be proposed.

The study that depicted the coincidence of the solutions computed by the algorithm Peirce to those derived through abductive reasoning performed by humans was not exhaustive because there is such a high number of domains, and it did not include the diversity and quantity of individuals. However, these results provided value suggesting that the abductive reasoning conducted by humans tends to coincide with the solutions computed by the algorithm Peirce.

References

1. Alberti, M., Gavanelli, M., Lamma, E.: The CHR-based Implementation of the SCIFF Abductive System. Fundamenta Informaticae 124 (4), pp. 365–381 (2013)
2. Aliseda, A.: Abductive Reasoning: Logical Investigations into Discovery and Explanation. Springer, Netherlands (2006)
3. Caroprese, L., Trubitsyna, I., Truszczynski, M., Zumpano, E.: A Measure of Arbitrariness in Abductive Explanations. To appear in Theory and Practice of Logic Programming, 25 p (2014)
4. Christiansen, H.: Executable specifications for hypothesis-based reasoning with Prolog and Constraint Handling Rules. Journal of Applied Logic 7 (3), pp. 341–362 (2008)
5. Dillig, I., Dillig, T.: Explain: A Tool for Performing Abductive Inference. In Sharygina, N., Veith, H. (eds) CAV 2013. LNCS, vol. 8044, pp. 684—689. Springer, Heidelberg (2013)
6. Eiter, T., Gottlob, G.: The Complexity of Logic-Based Abduction. Journal of the Association for Computing Machinery, 42 (1), pp. 3–42 (1995)
7. Fortier, N., Sheppard, J., Strasser, S.: Abductive inference in Bayesian networks using distributed overlapping swarm intelligence. Soft Computing Journal, May 2014, pp. 1–21 (2014)
8. Harman, G. H.: The inference to the best explanation. The Philosophical Review 74 (1), pp. 88–95 (1965)
9. Howard, P.: Introduction to Logic: Propositional Logic. Prentice-Hall, New Jersey (1999)
10. Josephson, J. R., Josephson, S. G.: Abductive Inference: Computation, Philosophy, Technology. Cambridge University Press, Cambridge (1994)
11. Kakas, A. C., Kowalski, R. A., Toni, F.: Abductive Logic Programming. Journal of Logic and Computation 2 (6), pp. 719–770 (1995)
12. Mackenzie, I. S.: Human-Computer Interaction: An Empirical Research Perspective. Morgan Kaufmann, New York (2013)
13. Magnani, L.: Abductive Cognition: The Epistemological and Eco-cognitive Dimensions of Hypothetical Reasoning. Springer, Berlin (2009)
14. Peirce, C. S.: Collected Papers of Charles Sanders Peirce. Oxford University Press, London (1958)
15. Romdhane, L. B., Ayeb, B.: An Evolutionary Algorithm for Abductive Reasoning. Journal of Experimental & Theoretical Artificial Intelligence 23, pp. 529–544 (2011)

16. Thagard, P. R.: The Best Explanation: Criteria for Theory Choice. The journal of philosophy 75 (2), pp. 76–92 (1978)
17. Thagard, P. R.: Explanatory Coherence. Behavioral and Brain Sciences 12, pp. 435–502 (1989)