Open Research Online

The Open University's repository of research publications
and other research outputs

An Ontology Design Pattern to Define Explanations

Conference or Workshop Item

oro.open.ac.uk

An Ontology Design Pattern to Define Explanations

Ilaria Tiddi, Mathieu d'Aquin, Enrico Motta
Knowledge Media Institute, The Open University, United Kingdom
{ilaria.tiddi, mathieu.daquin, enrico.motta}@open.ac.uk

ABSTRACT

In this paper, we propose an ontology design pattern for the concept of "explanation". The motivation behind this work comes from our research, which focuses on automatically identifying explanations for data patterns. If we want to produce explanations from data agnostically from the application domain, we first need a formal definition of what an explanation is, i.e. which are its components, their roles or their interactions. We analysed and surveyed works from the disciplines grouped under the name of Cognitive Sciences, with the aim of identifying differences and commonalities in the way their researchers intend the concept of explanation. We then produced not only an ontology design pattern to model it, but also the instantiations of this in each of the analysed disciplines. Besides those contributions, the paper presents how the proposed ontology design pattern can be used to analyse the validity of the explanations produced by our, and other, frameworks.

Categories and Subject Descriptors

I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods — *Ontology Design Patterns*

Keywords

Explanation, Ontology Design Pattern, Knowledge Discovery

1. INTRODUCTION

> "*The word explanation occurs so continually and holds so important a place in philosophy, that a little time spent in fixing the meaning of it will be profitably employed.*"
> John Stuart Mill – A System of Logic, 1843.

In this paper, we present an ontology design pattern to support the formal representation of an *explanation*. The motivation behind our work comes from our research, whose aim is to automatically find an explanation to data patterns using background knowledge from the Web of Data. Note that a data pattern is intended as a subset of data

behaving in the same regular way. Our general research problem is that finding an explanation for those patterns is an intensive task which is still relying on human experts, whose role is to provide explanations and refine results using their own background knowledge, while the cross-domain and machine-accessible knowledge of the Web of Data could be used to facilitate the automatisation of explaining data patterns.

The challenge we face here is that if we wish to give explanations to patterns or, more specifically, to process data in order to produce them, we first need a formal way to define what is an explanation $\mathcal{E}$[1]. Defining $\mathcal{E}$ is a complicated epistemic matter upon which a common agreement has never been reached, although this has been extensively discussed through time and across disciplines. Rather than entering this discussion, our proposition is to identify an ontology design pattern that formally supports the representation of an explanation – including its components, their roles and interactions – in order to provide an abstract description which can be applicable to any context where a system automatically produces explanations.

The methodology we used to derive such a design pattern is to explore how explanations are defined in those areas that most deal with the organisation and understanding of knowledge, usually grouped and defined as the "cognitive sciences". We reviewed the main literature in the disciplines embraced by Cognitive Science in order to see how their researchers see explanations. By identifying differences and/or commonalities among different areas, such as which aspects of an explanation matter and how $\mathcal{E}$ is defined, we aimed at abstracting our own definition for $\mathcal{E}$.

What we finally propose is a formal definition of $\mathcal{E}$ encoded as an ontology design pattern (ODP), presented with its instantiations into ontological representations of explanations in each of the considered disciplines of Cognitive Science. Those constitute, along with a cross-domain survey on the definition of explanation, the major contributions of our paper. Finally, we also show how to use the pattern to assess the validity of a framework producing explanations in real-world scenarios.

2. MOTIVATION

In our previous work [37], we presented Dedalo as a framework to explain groups of items using background knowledge extracted from Linked Data[2]. Dedalo's assumption is that

---

[1]We will refer to $\mathcal{E}$ further on, to intend "the concept of explanation".

[2]http://www.w3.org/standards/semanticweb/data

items appearing in the same group share underlying commonalities beside the dimensions used for their grouping, and these commonalities can be detected in the graph of Linked Data in the form of chains of RDF properties and a value shared by a large part of the items in the group to explain. We call those chains "candidate explanations". Because Linked Data can be traversed using URI dereferencing, Dedalo iteratively expands the graph using an A* strategy trying to find new, more accurate chains (evaluated with an ad-hoc F-Measure). Finally, a Linked Data pathfinding approach is used to establish the relation between the extracted candidate explanations and the criteria that grouped the items, in order to identify the most plausible explanations while removing noisy results.

In a very short example, we might want to understand why people search for the actor Daniel Radcliffe only at specific times[3]. The explanation, if being familiar with the context, can be identified as "people search for Daniel Radcliffe when a Harry Potter movie is released". While a human uses his own background knowledge to derive this (therefore, explaining the phenomenon), Dedalo uses the one from Linked Data. The problem here is that, if the human is not familiar with the context of Harry Potter, he will not find the above statement as a plausible explanation. Statistically computing explanations on the Linked Data graph has also its drawbacks, especially because some coincidences could emerge in the set of candidate explanations (e.g., "people search for Daniel Radcliffe when the U-20 Football World Cup is on"[4]).

From the example above, we understand that an explanation does not only mean putting two events in a correlation ("X happened because of Y"), but also defining the causal dependence based on the context in which they occur ("X happened because of Y in some context C"). The question is then whether this is enough to define an explanation and, if not, what else is needed for $\mathcal{E}$ to be complete. In a broader sense, in this work we aim at identifying which are the important actors of an explanation, i.e. which elements and interactions are needed to let us declare that, given a pattern and the derived observation, we have raised them to an explanatory level, not only to the one of a pure correlation. Our challenge consists then in designing an ontology pattern to formally represent it. In order to do so, we explored how the concept of explanation is defined within the disciplines embraced in Cognitive Science (in Section 3) to find their commonalities and differences in the view of an explanation and finally extract and apply (in Section 4) an ontology design pattern to define $\mathcal{E}$.

3. DEFINING EXPLANATIONS: A SURVEY

Cognitive Science was firstly defined by [16] as the science of understanding the nature of the human mind. In this work, dated 1985, Cognitive Science is presented as a field involving a series of disciplines interacting in such a way that they can be represented as a "cognitive hexagon" (Figure 1). The next sections provide a summary of how the main trends in these disciplines have tried to define explanations. It is worth mentioning that this is an organi-

---

[3]See the search peaks shown by the Google Trends tool, at http://www.google.com/trends/explore#q=Daniel%20Radcliffe
[4]The release of some Harry Potter movies and more than one U-20 football World Cup have indeed happened at the same time.
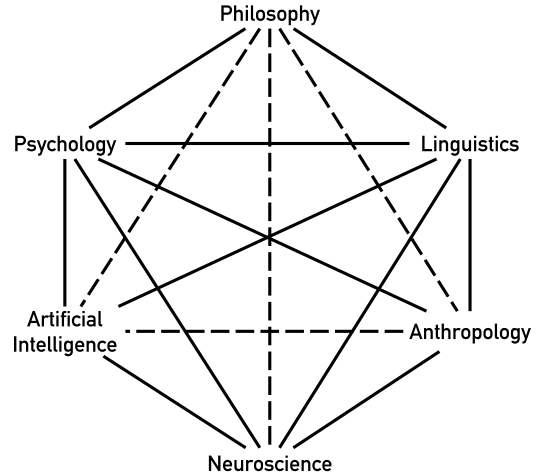

Figure 1: The cognitive hexagon as in [16]. Dotted lines are weaker connections between disciplines.

sational choice of our methodology, but many of the cited works, especially the contemporary ones, span over multiple disciplines. Additionally, we chose to consider the broader and more common fields of Computer Science and Sociology rather than the original Artificial Intelligence and Anthropology. The common nomenclature, that we also adopt, is to define *explanandum/nda* ("that which is explained") the fact or event to be explained and *explanans/tia* ("that which does the explaining") some premise to the explanandum.

3.1 Explanation and Philosophy

The works of [28, 30, 36] provide interesting surveys on how explanations have been defined by intellectuals from the ancient times – the first attempts of discussion already appear among the Greek intellectuals – to contemporary days. Thucydides (in his *History of the Peloponnesian War*) defines explanations as a process where facts, or "indisputable data", are observed, evaluated based on our knowledge of human nature and then compared in order to reach generalised principles for why some events occur. For Plato (in *Phedus* and *Theaetetus*), an explanation is an expression of knowledge using the *logos*, and is composed by the Forms, consisting in the abstractions of the entities we know ("the unchanging and unseen world") and the Facts, i.e. occurrences or states of affairs, which are the changing world we are used to see. Aristotle in his *Posterior Analytics* presents the explanation as a deduction process of finding out the cause of why something happened (according to Aristotle's 4-cases doctrine, the cause can be either the thing's matter, form, end, or change-initiator). In the modern age of determinism, causality becomes prominent to explanations. To know what caused an event means to know why it happened, which in turns means understanding the universe determined by natural laws. Descartes and Leibniz describe explanations as the process of demonstrating the mechanical interactions between God ("the primary efficient cause of all things"), the world things (secondary causes), the laws of nature and some initial conditions. Newton rejects the God component and puts the laws of nature as playing the central role. For him, explanation of natural phenomena is a deductive process which finds the most general principles (the fundamen-

tal laws of nature) that account for them. The Empiricists then reject explanations which cannot be traced directly to experience. According to Hume, all causal knowledge stems from experience, and causation is just a regular succession of facts to be discovered by experience. Kant, reintroducing the metaphysical component into explanations, believes that the knowledge starts with experience, but does not arise from it: It is shaped by the categories of the understanding and the forms of pure intuition (space and time). John Stuart Mill rejects this view and defines $\mathcal{E}$ as only based on the laws of nature and on deduction. A fact is explained by deducing its cause, that is, by stating the laws of which it is an instance. Carl Hempel, who revisited and extended Mill's work, started the contemporary discussions around explanations. Explaining an event is to show how this event would have been expected to happen, taking into account the laws that govern its occurrence, as well as certain initial conditions. Formally speaking, a singular explanandum E is explained if and only if a description of E is the conclusion of a valid deductive argument, whose explanantia involve essentially a set C of initial or antecedent conditions and a law-like statement L [20]. L can be either statistical (in inductive-statistical explanations) or laws of nature (in deductive-nomological explanations). Salmon [32] further extends the inductive-statistical explanation model proposing the statistical-relevance model, where explanations have to capture the dependencies between the explanandum and the explanans statistically, and the causal model, which sees explanations as mechanical processes of interacting components. An interactive view is also the one of the Interventionists [39], where explanations are the process of manipulating the explanandum within a space of alternative possibilities to see if a relation holds when various other conditions change. The Unificationism approach of Friedman and Kitcher [13, 23] affirms that explaining a phenomenon is to see it as an instance of a broad pattern of similar phenomena. According to Friedman, explanations are based on both the scientific understanding of the world, but also on some assumptions (i.e. regularities in nature) which are the basic "unifiers". Similarly, Kitcher believes that a number of explanatory patterns (or schemata) need to be unified in order to explain a fact.

3.2 Explanation and Psychology

Explanations in psychology are generally meant to regulate behaviours or to process some information, i.e. they are "an attempt to understand phenomena related to intelligent behaviours" [2]. Many works of psychology criticise the traditional philosophical-scientific approaches to explanation since they are too much focused on causality and subsumption under laws (deduction) [2, 8, 21]. Instead, psychological explanations accept over-determinism and dynamism, i.e. the antecedent conditions that explain a phenomenon can be more than one and change over time. The explanantia are not laws of nature but human capacities, that psychological methodologies have to discover and confirm. In the literature, psychological explanations have been defined as "models" providing a law, by means of which it is possible to describe a phenomenon while relating it to other, similar phenomena, and possibly to allow a prediction, intended as a future occurrence of the phenomenon [8, 21]. Authors of [26] have highlighted how the mechanical aspect within the explanation process plays a central role into the

recent psychology works. Phenomena are often explained by decomposing them into operations localised in parts (components) of the mechanism. Mechanisms have spatial and temporal organisations that explain how entities are organised into levels and carry out their activities. In other words, explaining a phenomenon means revealing its internal structure by defining the organisational levels which are responsible of producing it, then identifying how those levels relate to other levels, and finally building a model to explain similar phenomena. This idea of interactive levels can be found in the major contemporary psychology trends, i.e. the connectionist (bottom-up) approach [2, 9, 25], that first specifies the system's components and then produces the explanation's model, and in the symbolic (top-down) one [34], that first identifies the phenomenon to be explained and then describes it according to the external symbols (syntactico-semantic rules or mathematical laws) manipulated by the humans.

3.3 Explanation and Neuroscience

Neuroscience explanations reject the concept of levels, as well as the synthetic a priori frameworks. Mentality is the central pivot between interpretation and rationality. Neuroscientists aim at understanding minds, and this requires to assume that beliefs are true and rational, which is in contrast with many philosophical views. Explaining, in neuroscience, means fitting a mental phenomenon into a broadly rational pattern. Explanations are answers to the what-if-things-had-been-different questions: they describe what will happen to some variables (effects) when one manipulates or intervenes on some others (causes) [4, 10, 40]. They can be "upper level explanations", when causes are macroscopic variables or environmental factors, and "lower level explanations", if causes are more fine-grained (neural, genetic, or biochemical) mechanisms. Also, explanations do not require any law nor sufficient condition, and are simply generalisations that describe relationships stable under some appropriate range of interventions. Therefore, a typical approach to explanations consists in detecting how the variables respond to the hypothetical experiments in which some interventions occur, then establishing if other variables under the same intervention respond in the same way. The relevance of explanations is assessed by their stability, i.e. some relationships are more stable than others under some interventions or under some background conditions.

3.4 Explanation and Computer Science

In Artificial Intelligence (AI), explanations are considered in terms of inference and reasoning. Programs are meant to decipher connections between events, in order to predict other events and represent some order in the universe. The process of generating explanations fits the two reasoning patterns, abduction and induction, that Peirce describes in its *Lectures on Pragmatism* of 1903. According to Peirce, induction is the inference of a rule (major premise) starting from a case (minor premise) and its result (conclusion), while abduction is the inference of the case from a rule and a result [12]. Those models have been often put under the same umbrella-term of non-deductive, a posteriori inference, and the same Peirce in its later period sees abduction as only one phase of the process, in which hypotheses and ideas are generated, but then need to be evaluated by induction or deduction [12, 31, 33]. Moving on from such discus-

sion, AI approaches to generate explanations (see [27, 29, 31] for extensive surveys) show that the two inferences can be similarly modelled, so that one body of initial knowledge (divided into observations and prior knowledge) is mapped to another (the new knowledge), under constraints of certain criteria, which reveal their connection and confirm the hypothesis. The explanation process then consists in first analysing the observations, then defining tentative hypotheses and finally proving them empirically. Similarly to explanations in psychology, Artificial Intelligence explanations do not assume that a truth has been antecedently accepted in the prior knowledge: In order to produce new knowledge, they need to be scientifically and empirically evaluated. Plausibility, usefulness and desirability are considered too subjective and are replaced by some formal constraints that the process has to respect, namely, efficiency (how "easily" it solves the problem), non-deducibility (it should introduce new knowledge), consistency (it must be consistent with what it is already known) and non-emptiness (it must introduce new knowledge). Some works attempting to merge the views from philosophy and Artificial Intelligence are surveyed in [27]. Explanations also play a central role in other areas of Computer Science such as Knowledge Management. Knowledge Management systems are used for intelligent tutoring or decision-support, and explanations provide insight into how their reasoning is accomplished. Users are presented solutions and explanations of a problem to be solved relying on some domain knowledge (see referenced works in [1, 35, 41]). This domain knowledge, related to facts, procedures or events, is defined as "tacit" because it is already in the individuals' mind. An inflow of new stimuli then triggers a cognitive process which produces explanations, which are the explicit knowledge communicated in symbolic forms. Explanation can be intended as "giving something a meaning or an interpretation to, or to make it understandable" [41].

3.5 Explanation and Sociology

For Weber and Durkheim [11, 38], founders of the major two trends in sociology, explaining is the act of giving a meaning and justifying social facts, where those are intended as observable regularities in the behaviour of the members of a society. In their perspective, explanation is a rational (empirical) observation of social behaviours. More contemporary works extend this view, by defining explanations as a comparison of events [5] or as a form of reasoning [3], which have to take in account the social world and social behaviours constantly evolving in time and space. Generally, the agreement is that explanations, based on social facts, have more pragmatic requirements with respect to scientific explanations. Formal constraints, such as mathematical formalisation or empirical falsifiability, leave space to descriptions and approximations of complex structures of the social world [6, 17]. Explanations might be weaker or stronger depending on the set of social impacts that the rules express. For example, they can be proverbs (everyday rules expressing what one knows), social theorems (rules summing up the human experience) or paradoxes [3, 24]. Some have also distinguished between social-anthropological and religious explanations, where "religious" means giving sense to phenomena using symbols of a tradition. Explanations are declared to be valid if a phenomenon can also appear due to other circumstances, and logically exhaustive if they are generalisable to all logically possible combinations of conditions encountered [3, 24].

3.6 Explanation and Linguistics

Linguistic explanations are descriptions focused on finding what rules of a language explain a linguistic fact [19]. Linguists do not aim at describing the processes but rather at explaining how language acquisition is possible, by identifying the grammatical patterns that occur in human languages. Useful surveys on explanations in linguistics are to be found in [7, 18, 19, 22]. A central assumption in linguistic theories is that a language is a set of elements (words, morphemes, phonemes) that meet a set of wellformedness conditions, called language universals. Those consist in language-specific characteristics that form the grammar of the language. The major approaches, the generative linguistics (following Chomsky's theories) and the functional linguistics (based on Greenberg's school) have different views on the roles and functions of the components of linguistic explanations. In the former case, only the generalisation of an explanation is regarded as interesting, since it shows that linguistic constructions are derivable from the general regularities of a language (*descriptive adequacy*), but some of those regularities are innate and have to be accepted a priori (*explanatory adequacy*). The generativists see the act of explaining as defining new constraints on the frameworks which are able to describe the languages. This claim to innateness and the view of explanations as constrained descriptions cannot be accepted by functional linguistics, for which external elements and realities are essential to understand linguistic facts. Functionalists are more interested in language regularities than in descriptive frameworks. The rejection of a priori language universals does not mean that in language structure is arbitrary, but rather than it is worth attempting to explain, wherever possible, language facts using external non-linguistic factors [18].

3.7 Summary Considerations

From the sections above, we briefly summarise the following conclusions.
- Philosophy is the "discipline of asking questions and checking answers" [16]. An explanation in Philosophy is a process where a set of initial elements (an event and some initial conditions) are described, deduced or put into a relation with an output phenomenon according to a set of (empirical or metaphysical) laws.

- Psychology tries to understand the human (and animal) cognitive processes. Here, $\mathcal{E}$ is intended as a model of components (entities and activities) that mechanically interact to produce regular changes from some starting conditions to a termination condition.

- Neuroscience gives an understanding of the brain. Explanations are seen as generalisations that describe relationships between input and output variables, that are stable under some appropriate range of interventions without any law nor sufficient condition.

- Computer Science deals with computer programs as a medium to perform human operations. Explanations are generalisations of some starting knowledge (some facts and some prior knowledge) mapped to another, new knowledge under constraints of certain criteria.

- Sociology focuses on the links between the human's cognitive processes and his sociocultural world. It intends
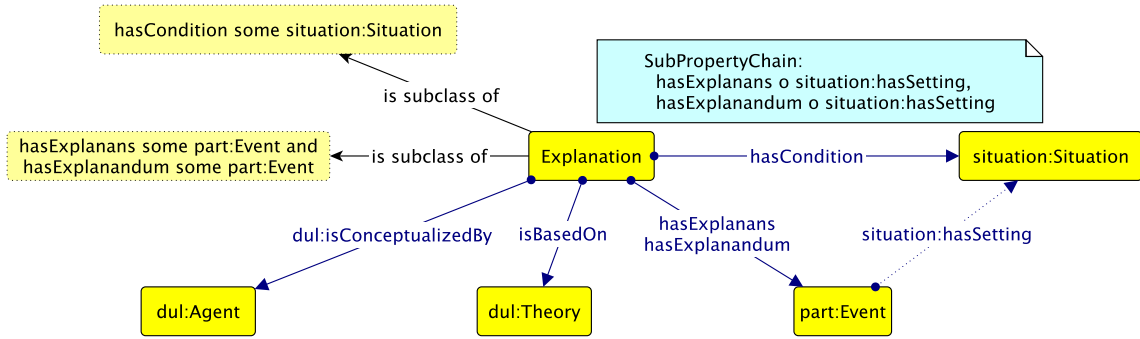
Figure 2: The Explanation Ontology Design Pattern.

explanations as the act of giving a meaning to some social regularities occurring because of some social behaviours.

- Linguistics provides theories dealing with the nature of the human language. An explanation is the process of deriving the regularities of a language (the grammar) from a set of linguistic-specific facts respecting some wellformedness constraints.

With those premises, it is possible to abstract a common model for $\mathcal{E}$. The linguistic formalisation we found particularly suitable for our purposes was the one of [24] and presented hereafter:

When X happens, then, due to a given set of circumstances C, Y will occur because of a given law L.

The next section is focused on the design of the ontology pattern representing $\mathcal{E}$, as well as its instantiations in the various presented disciplines.

4. THE EXPLANATION DESIGN PATTERN

Ontology Design Patterns are small, motivated ontologies to be exploited as building blocks in ontology design [15]. Their purpose is to be used as designing components of a problem with the assumption that an ontology should be modelled according to those patterns, with appropriate dependencies between them, plus some necessary design expansion based on the specific situation's needs. We followed this idea to build an ontology design pattern that represents an explanation. We reused some of the existing design patterns and ontologies, and then extended classes and properties according to our needs. The resulting model, that we call $\mathcal{E}$-ODP, is shown in Figure 2 and described hereafter[5].

4.1 Ontology Design Pattern

Shortly, we define $\mathcal{E}$-ODP as an ontology for characterising what is an explanation to an event, and include the various components that make the process plausible (for instance, the initial conditions under which the explanation happens, the outcome event, and so forth).

(a) The main class *Explanation* represents $\mathcal{E}$, which we aim to identify. It will be differently instantiated according to the disciplines we have presented. The class has few constraints that we represent as OWL class restrictions. In order to be complete, an explanation: (i) needs at least one antecedent event, expressed as *hasExplanans* `some` *part:Event*; (ii) requires a posterior event,

expressed as *hasExplanandum* `some` *part:Event* and (iii) has to happen in a context that relates the two events, expressed as *hasCondition* `some` *situation:Situation*.

(b) An explanation has then an antecedent event (the explanans) and a posterior event (the explanandum). We saw how those have been called not only events but also phenomena, facts, variables, observations, knowledge, etc. We chose to reuse the *part:Event* class from the Participation ODP[6], which enables to represent any binary relation between objects and events. Thanks to this, our pattern can be easily integrated in the Participation one by extending the *part:Event* individuals according to the structure of the Participation ODP.

(c) Two events have to be related in context, otherwise it is not possible to distinguish their correlation from a coincidence. To represent the conditions (or constraints) under which the two events occur, we used the *situation:Situation* class from the Situation ODP[7], which is precisely designed to "represent contexts or situations, and the things that have something in common, or are associated" (see documentation). The OWL axiom in the blue box above the object property *hasCondition* is a Manchester Syntax axiom showing what we can infer about the situation in which the explanation is happening. If there exists a situation in which an explanation is contextualised (through *hasCondition*), then both the explanans and the explanandum do share this same situation (through *situation:hasSetting*), therefore they are in the same context.

(d) The last main component of an explanation is represented using the *dul:Theory* class from the DOLCE+DnS Ultralite ontology[8]. According to the documentation, "a theory is a description that represents a set of assumptions for describing something, usually general. Scientific, philosophical, and common-sense theories can be included here". We use the term theory with the idea that it best embraces those concepts defined in Cognitive Science (laws, human capacities, human experiences, universals, and so on), all consisting in something having a binding force on some events under analysis. Note that the Theory is a specialisation of the Description class from the Descriptions&Situations ontology [14]: In this view, our theory acts as the description that classifies the explanation (corresponding to the

[6]http://ontologydesignpatterns.org/cp/owl/participation.owl
[7]http://www.ontologydesignpatterns.org/cp/owl/situation.owl
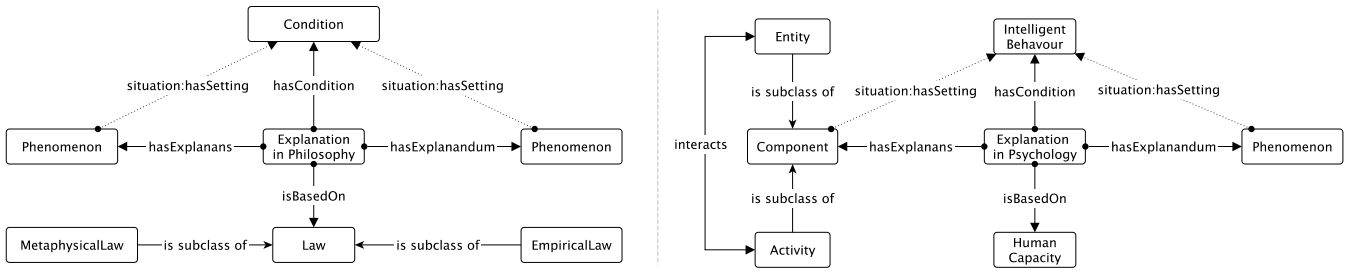[8]http://www.ontologydesignpatterns.org/ont/dul/DUL.owl

Figure 3: Instances of the pattern in Philosophy (left) and Psychology (right).

situation), which is built upon the context, the event to be explained, and the possible explaining events.

(e) We also include the DOLCE class *dul:Agent* to represent the one producing the explanations. Since there is no agreement on how to "call" the act of explaining (e.g., deducing, inducing, inferring, describing have all been used in the literature), we then kept the generic object property label as *dul:isConceptualizedBy*.

4.2 Instances of the Pattern

Finally, we create instances of the $\mathcal{E}$-ODP according to our findings in the survey of Section 3. We present the ontological model for each discipline as well as a practical example which gives an idea of how the process works. In the examples, we use the notation $\mathcal{E}{=}(\mathcal{A}, \mathcal{P}, \mathcal{C}, \mathcal{T})$ where $\mathcal{A}$ stands for the antecedent event/explanans, $\mathcal{P}$ for the posterior event/explanandum, $\mathcal{C}$ for the situational context they are happening in and $\mathcal{T}$ for the theory governing those events. For the purpose of avoiding redundancy, the class *dul:Agent* is omitted. Note that all the presented patterns have been submitted to the ODP central hub[9] but are also available online with a higher resolution[10].

Philosophy. The pattern for $\mathcal{E}$ as seen in Philosophy is presented in Figure 3 (left). $\mathcal{T}$ is a *Law*, that we also sub-specified as metaphysical or empirical, trying to sketch some of the general views in the area (roughly, pre- and post-empirical views). Both $\mathcal{A}$ and $\mathcal{P}$ are *Phenomenon* classes while $\mathcal{C}$ is a necessary and sufficient *Condition* class. It is worth mentioning that ours is a mere generalisation of how an explanation is thought in Philosophy, and that this can be in turn specialised into more specific patterns according to, for instance, different times or trends. For example, we could represent Mill's explanation model as $Mill_{\mathcal{E}}{=}$ (*CausingFact, EffectFact, InitialCondition, LawOfNature*) while Plato's model as $Plato_{\mathcal{E}}{=}$ (*InitialFact, PosteriorFact, Circumstance, Form*).

Psychology. The pattern for $\mathcal{E}$ as intended in Psychology is presented in Figure 3 (right). We have shown in Section 3 how psychological explanations are more focused on their mechanical aspects. Therefore, $\mathcal{A}$ is an *InitialComponent* class, whose subclasses are the interacting classes *Entity* and *Activity*. The context $\mathcal{C}$ is the human *IntelligentBehaviour* that correlates $\mathcal{A}$ and $\mathcal{P}$. The theory consists in a *HumanCapacity* and not in laws of nature. In a practical example, a simple event $\mathcal{P}$ as "slipping on a banana" can be explained by the fact $\mathcal{A}$ that the person did not to pay attention to the path. A psychological explanation would attempt to relate

the two events to ($\mathcal{C}$) the mental state and feelings of the person slipping as, for instance, anger or tiredness. If this condition is not met anymore (e.g., one is calm and careful), the explanation does not hold. The human capacities which govern the examples are the ones ruling the human's mind so that the human actions change depending on the feelings.

The remaining patterns are shown in Figure 4 in a clockwise order.

Neuroscience. The idea of explanations in Neuroscience is similar to the one in Psychology, with the difference that the events are considered as *CauseVariable* ($\mathcal{A}$) and *EffectVariable* ($\mathcal{P}$). We have seen how neuroscientists reach explanations by *Experiment*(s) ($\mathcal{T}$) rather than with laws, and that the necessary conditions $\mathcal{C}$ are rather called *Intervention*(s) under which the two events occur. If those are changed, they might nullify the explanation. For example, we explain why humans can do math calculations ($\mathcal{P}$) by the fact $\mathcal{A}$ that some particular neurons respond actively to quantities. The neuroimaging has proven that ($\mathcal{T}$) some areas on the parietal lobe are used to provide pre-verbal representations. The context $\mathcal{C}$ in which the explanation holds is the one of numerical calculations.

Computer Science. The model for an explanation in Computer Science can be easily represented as $CS_{\mathcal{E}}{=}(Observation, NewKnoweldge, Constraint, PriorKnowledge)$. Note that the new knowledge can be *GeneralisedNewKnowledge*, if the explanation is reached by induction, or *SpecialisedNewKnowledge*, if reached by abduction. We use here the example given by [31]: a caveman is roasting a lizard ($\mathcal{C}$) on the end of a pointed stick and he is watched by an amazed crowd of cavemen, who have been using for years only their bare hands. The cavemen explain that painless roasting ($\mathcal{P}$) is effect of ($\mathcal{A}$) using a long, rigid, sharp object simply by observing (which is equivalent to using the prior knowledge $\mathcal{T}$) how the stick supports the lizard while keeping the hand away from the fire.

Sociology. Sociology reintroduces the human component into its idea of explanation. We represent it as $S_{\mathcal{E}}{=}(SocialBehaviour, SocialRegularity, SocialWorld, HumanExperience)$. The social trend $\mathcal{P}$ that Italians live with their parents until later ages is explained by ($\mathcal{A}$) the constantly decreasing job opportunities, which makes impossible living on its own ($\mathcal{C}$). The theory behind it are the rules governing the socio-economical behaviours.

Linguistics. Explanations in Linguistics are models focusing on the function and nature of the language, namely a series of *LinguisticFact*(s) specific to a *Language* are analysed on the basis of some *LanguageUniversal*(s) to finally

---

[9]http://ontologydesignpatterns.org/
[10]http://people.kmi.open.ac.uk/ilaria/experiments/explain_odp/

Intervention

situation:hasSetting — hasCondition — situation:hasSetting

Cause Variable — hasExplanans — Explanation in Neuroscience — hasExplanandum — Effect Variable

isBasedOn

Experiment

Constraint

situation:hasSetting — hasCondition — situation:hasSetting

Observation — hasExplanans — Explanation in Computer Science — hasExplanandum — NewKnowledge

isBasedOn

Prior Knowledge

Language

situation:hasSetting — hasCondition — situation:hasSetting

Linguistic Fact — hasExplanans — Explanation in Linguistics — hasExplanandum — Grammar

isBasedOn

Language Universal

SocialWorld

situation:hasSetting — hasCondition — situation:hasSetting

Social Behaviour — hasExplanans — Explanation in Sociology — hasExplanandum — Social Regularity
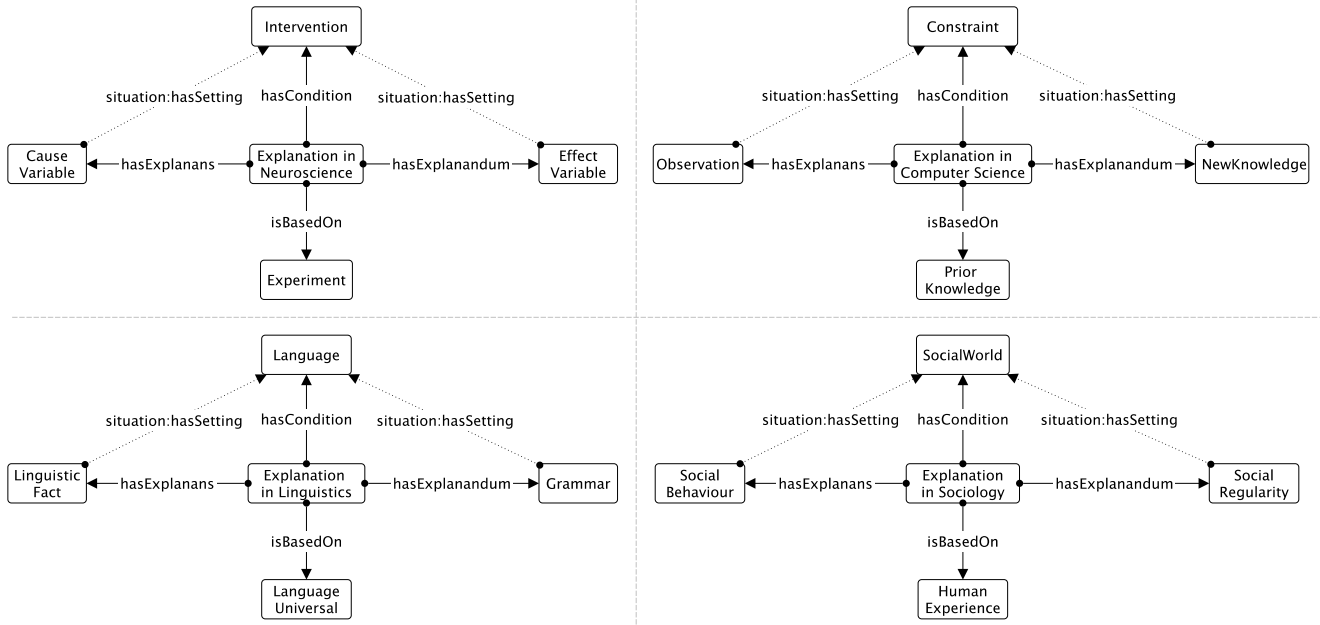
isBasedOn

Human Experience

Figure 4: Instances of the pattern for (clockwise) Neuroscience, Computer Science, Sociology and Linguistics.

derive a *Grammar*. For example, to explain that English ($\mathcal{C}$) does not allow the expression "**the my book*" ($\mathcal{P}$), one has to know that both *my* and *the* are determiners ($\mathcal{A}$), and that ($\mathcal{T}$) only one determiner is accepted.

4.3 Pattern Application on Real Frameworks

Once we have derived the pattern for $\mathcal{E}$, we can return to our original motivating scenario and analyse how much Dedalo's framework fits into the process of deriving an explanation. The same approach can be of course applied to any other framework that automatically produces explanations. The scenario, already presented in Section 2, is the following: Dedalo wants to explain the fact $\mathcal{P}$ that a group of users is searching the Web for information about Daniel Radcliffe only at some very specific times. What we need to identify is what constitutes $\mathcal{A}$, $\mathcal{T}$ and $\mathcal{C}$.

The first component of Dedalo uses the Linked Data background knowledge to derive some plausible explanantia of type $\mathcal{A}$ to the event $\mathcal{P}$. The process consists in finding facts in Linked Data that are statistically highly correlated with the explanandum. For example, both "people search for Daniel Radcliffe because a Harry Potter movie is released" and "people search for Daniel Radcliffe because the Under-20 Football World Cup is on" are correlated with Daniel Radcliffe according to Linked Data information[11]. If no connection between the $\mathcal{A}$ events and $\mathcal{P}$ is identified, however, there is no way to distinguish whether the $\mathcal{A}$ and $\mathcal{P}$ are happening together by coincidence (as in the second case). Also, the explanation is incomplete (and impossible) if one has no knowledge of the context $\mathcal{C}$ (what is the relation between Daniel Radcliffe and Harry Potter).

Dedalo's second component is therefore the one finding this relationship, in the form of a Linked Data path between two entities representing the facts (for instance, Daniel Radcliffe and Harry Potter/the U-20 Football World Cup) to

provide a more explicit information to the user. For example, there exists a Linked Data path such as $\langle$`DanielRadcliffe portrays HarryPotter`$\rangle$ but not between Daniel Radcliffe and the U-20 Football World Cup. This path explicits the context $\mathcal{C}$ in which those two facts are happening: without it, the two events are nothing but a spurious correlation.

This brings us to our conclusion point. The obtained explanation can be summarised as "people search for Daniel Radcliffe because a Harry Potter movie is released, since Daniel Radcliffe is the main character in it". A familiar reader might recognise here the law $\mathcal{T}$ governing the explanation: actors play in movies, therefore it makes sense that people are interested in Daniel Radcliffe if he is the one playing in Harry Potter movies. Dedalo is, however not (yet) able to provide the same information: the user to whom the explanation is presented is left inferring that Daniel Radcliffe is an actor, that "Harry Potter" is a movie and, equally, that "actors play in movies" (as a general principle).

5. CONCLUSIONS

This work is an attempt to formally define what is intended as an "explanation". We have presented such formalisation as an ontology design pattern, whose modelling has been achieved by analysing how the different disciplines of Cognitive Science perceive explanations, i.e. which elements and interactions are needed for an explanation to be in place. The proposed ontology pattern has then been instantiated according to the surveyed disciplines. Finally, we have shown how to apply the pattern on custom frameworks that automatically produce explanations in real-world scenarios, by showing how to assess whether the system fits the proposed pattern.

We identified as an important future work the introduction in our Linked Data-based framework of a third component, whose role is to explicit the laws underlying the relationship between the events, so that the explanation pre-

---

[11]To test this scenario, see http://linkedu.eu/dedalo/demo

sented to the users can be considered complete according to the proposed $\mathcal{E}$-ODP. Another direction to explore is the possibility of refining the pattern by introducing which qualities are necessary to an explanation (e.g., validity, truth, and so on) or by exploring new perspectives (e.g., other disciplines). Finally, an evaluation of other existing frameworks on the basis of our ontology design pattern is also possible.

6. REFERENCES

[1] M. Alavi and D. E. Leidner. Knowledge management systems: issues, challenges, and benefits. *Communications of the AIS*, 1(2es):1, 1999.

[2] W. Bechtel and C. Wright. What is psychological explanation. *Routledge companion to the philosophy of psychology*, pages 113–130, 2009.

[3] E. Brent, A. Thompson, and W. Vale. Sociology a computational approach to sociological explanations. *Social Science Computer Review*, 18(2):223–235, 2000.

[4] J. Campbell. Causation in psychiatry. *Philosophical issues in psychiatry*, pages 196–215, 2008.

[5] J. R. Carter. Description is not explanation: A methodology of comparison. *Method & Theory in the Study of Religion*, 10(2):133–148, 1998.

[6] P. Clayton. Explanation from physics to the philosophy of religion. *International journal for philosophy of religion*, 26(2):89–108, 1989.

[7] P. W. Culicover and R. Jackendoff. *Simpler syntax.* Oxford University Press Oxford, 2005.

[8] R. Cummins. How does it work?" versus" what are the laws?": Two conceptions of psychological explanation. *Explanation and cognition*, pages 117–144, 2000.

[9] L. Darden and N. Maull. Interfield theories. *Philosophy of Science*, pages 43–64, 1977.

[10] D. Davidson. The structure and content of truth. *The journal of philosophy*, pages 279–328, 1990.

[11] E. Durkheim. The rules ofthe sociological method, 1938.

[12] K. T. Fann. *Peirce's theory of abduction.* Springer Science & Business Media, 1970.

[13] M. Friedman. Explanation and scientific understanding. *The Journal of Philosophy*, pages 5–19, 1974.

[14] A. Gangemi and P. Mika. Understanding the semantic web through descriptions and situations. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, pages 689–706. Springer, 2003.

[15] A. Gangemi and V. Presutti. Ontology design patterns. In *Handbook on ontologies*, pages 221–243. Springer, 2009.

[16] H. Gardner. *The mind's new science: A history of the cognitive revolution.* Basic books, 2008.

[17] P. S. Gorski. The poverty of deductivism: A constructive realist model of sociological explanation. *Sociological Methodology*, 34(1):1–33, 2004.

[18] M. Haspelmath, H.-J. Bibiko, J. Hagen, and C. Schmidt. *The world atlas of language structures*, volume 1. Oxford University Press Oxford, 2005.

[19] J. A. Hawkins. *Explaining language universals.* Blackwell, 1988.

[20] C. G. Hempel and P. Oppenheim. Studies in the logic of explanation. *Philosophy of science*, pages 135–175, 1948.

[21] E. H. Hutten. On explanation in psychology and in physics. *The British Journal for the Philosophy of Science*, 7(25):73–85, 1956.

[22] E. Itkonen. On explanation in linguistics. In *article in the forum discussion of this issue: http://www. kabatek. de/energeia/E5discussion*, 2013.

[23] P. Kitcher. Explanatory unification. *Philosophy of Science*, pages 507–531, 1981.

[24] E. Maaløe. *Modes of Explanation: From Rules to Emergence.* Handelshøjskolen, Aarhus Universitet, Institut for Ledelse, 2007.

[25] D. Marr and A. Vision. A computational investigation into the human representation and processing of visual information. *WH San Francisco: Freeman and Company*, 1982.

[26] M. Marraffa and A. Paternoster. Functions, levels, and mechanisms: Explanation in cognitive science and its problems. *Theory & Psychology*, page 0959354312451958, 2012.

[27] A. Páez. Artificial explanations: the epistemological interpretation of explanation in ai. *Synthese*, 170(1):131–146, 2009.

[28] S. Psillos. Past and contemporary perspectives on explanation. *General philosophy of science. Focal issues*, pages 97–173, 2007.

[29] N. J. Rothleder. Induction: Inference and process, 1996.

[30] D.-H. Ruben. *Explaining explanation.* Cambridge Univ Press, 1990.

[31] S. Russell, P. Norvig, and A. Intelligence. Artificial intelligence. a modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, 25, 1995.

[32] W. C. Salmon. Scientific explanation: causation and unification. *Critica: Revista Hispanoamericana de Filosofia*, pages 3–23, 1990.

[33] G. Schurz. Patterns of abduction. *Synthese*, 164(2):201–234, 2008.

[34] P. Smolensky. On the proper treatment of connectionism. *Behavioral and brain sciences*, 11(01):1–23, 1988.

[35] R. W. Southwick. Explaining reasoning: an overview of explanation in knowledge-based systems. *The knowledge engineering review*, 6(01):1–19, 1991.

[36] M. Strevens. Scientific explanation. *Encyclopedia of Philosophy, second edition. DM Borchert (ed.). Detroit: Macmillan Reference USA*, 2006.

[37] I. Tiddi, M. d'Aquin, and E. Motta. Dedalo: Looking for clusters explanations in a labyrinth of linked data. In *The Semantic Web: Trends and Challenges*, pages 333–348. Springer, 2014.

[38] M. Weber and W. V. Heydebrand. *Sociological writings.* Continuum International Publishing Group Ltd., 1994.

[39] J. Woodward. *Making things happen: A theory of causal explanation.* Oxford University Press, 2003.

[40] J. Woodward. Cause and explanation in psychiatry. *Philosophical issues in psychiatry. Explanation, phenomenology, and nosology*, pages 132–184, 2008.

[41] B. A. Wooley. Explanation component of software system. *Crossroads*, 5(1):24–28, 1998.